Expanding the phenotype and genetic spectrum of myoclonic astatic epilepsy

Tang, Shan

Awarding institution:
King's College London

EXPANDING THE PHENOTYPE AND GENETIC SPECTRUM OF MYOCLONIC ASTATIC EPILEPSY

Dr. Shan Tang
MBBS, MRCPCH

Institute of Psychiatry, Psychology & Neuroscience
King's College London

A Thesis submitted for the Degree of Doctor of Philosophy to the University of London

February 2017

Abstract

Myoclonic astatic epilepsy (MAE) is a rare generalised childhood epilepsy with variable but poorly described neurodevelopmental outcome. Family studies suggest a major genetic influence as up to two thirds of relatives have seizures, or electroencephalographic (EEG) abnormalities. MAE is associated with 10 different genes, yet these genes account for less than 20% of the genetic aetiology of MAE leaving the majority unexplained.

The aims of this thesis were (1) describe the epilepsy and neurodevelopmental phenotype of MAE cases, (2) perform EEG studies on first degree family members for familial EEG abnormalities and compare occurrence of epileptiform features to population prevalence and (3) to collect DNA and identify MAE causative genetic variants through exome sequencing.

I assembled the largest MAE cohort (n=123) to date. The epilepsy phenotype is remarkably similar to previously published cohorts. I identified a severe neurodevelopmental phenotype: intellectual disability was reported in 64.9%, autism spectrum disorder in 21.3% and attention deficit hyperactivity symptoms in 41.0%. Additionally, extremely low adaptive behavioural scores were identified in 69.4% of cases. I performed EEG studies on 38 first-degree relatives of 13 MAE families, and found an excess of epileptiform EEG features in adults (>16 years), compared to controls (P=0.05, RR 6.82).

I identified likely pathogenic or candidate variants in 11 of 109 cases. This comprised known genes associated with MAE: *CHD2* n=1, *SYNGAP*1 n=2, *SLC6A1* n=1, *KIAA2022* n=1; epilepsy associated genes novel for MAE: *KCNB1* n=1, *MECP2* n=1, *KCNH5* n=1, and three new candidate genes; *SMARCA2* n=1*, ASH1L* n=1 and *CHD4* n=1.

Lastly, I highlight phenotypic features which help correlate with known and novel specific gene associations, discuss that MAE is a phenotypic and genetic nosological bridge between genetic generalised epilepsy and epileptic encephalopathy, and discussion applications and future directions leading on from this project.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would first like to thank my supervisors Professor Deb Pal and Professor Michael Simpson for their mentorship and availability throughout this project. I am particularly appreciative for their patience, flexibility and understanding of my journey through this PhD. I am also grateful to the College, Professor Mark Richardson, Professor Chris Shaw and the department for supporting my aspirations. I would like to thank the MRC for funding my Clinical Research Training Fellowship.

I am indebted to the talents of: Laura Addis, for her teaching in molecular genetics techniques and proofreading relevant chapters; Chayanin Tagwiriyasakil, for supporting quantitative EEG analysis; Michael Absoud, for training in neurodevelopmental assessments; Anna Smith, for reviewing the neuropsychological data and statistical analysis; Sushma Goyal and Stuart Smith for their teaching on EEGs; Simon Topp, Inti Pedrosa, Nick Dand and Christos Petridis for their expertise in bioinformatics.

I am grateful to the many UK and European clinical collaborators, research managers, research nurses, research assistants and EEG technicians who supported this project.

My thanks to the generosity and altruism of families who took part in the study, often undertaking long distance travel and sacrificing days off to take part.

Finally, special thanks to my family, to my parents for their best efforts in raising me, to my children for keeping my life and perspective in balance, and to my husband who supported me in every way he could.

Publications

Papers published either as a direct result from or through collaborative work during this thesis:

Tang S, Hughes E, Lascelles K, EuroEPINOMICS RES myoclonic astatic epilepsy working group, Simpson MA, Pal DK. New SMARCA2 mutation in a patient with Nicolaides-Baraitser syndrome and myoclonic astatic epilepsy. *Am J Med Genet A.* 2017 Jan;173(1):195-199.

Larsen J, Johannesen KM, Ek J, **Tang S**, Marini C, Blichfeldt S, Kibaek M, von Spiczak S, Weckhuysen S, Frangu M, Neubauer BA, Uldall P, Striano P, Zara F, MAE working group of EuroEPINOMICS RES Consortium, Kleiss R, Simpson M, Muhle H, Nikanorova M, Jepsen B, Tommerup N, Stephani U, Guerrini R, Duno M, Hjalgrim H, Pal D, Helbig I, Møller RS. The role of SLC2A1 mutations in myoclonic astatic epilepsy and absence epilepsy, and the estimated frequency of GLUT1 deficiency syndrome. *Epilepsia.* 2015 Dec;56(12):e203-8.

Declaration

The work presented in this thesis is my own work except where stated below.

Chapter 1

This chapter is my own work.

Chapter 2

Professor Deb Pal first established the Genetics of Human epilepsy study; which I amended to include myoclonic astatic epilepsy and elements of this project.

Chapter 3

This chapter is my own work.

Chapter 4

Dr Adam Pawley and Dr Chayanin Tangwiriyasakil wrote the quantitative EEG analysis matlab scripts.

Chapter 5

Neurophysiological technicians at King's College Hospital, St Thomas' Hospital and Evelina Children's Hospital performed EEGs. EEGs were reviewed jointly with Professor Michalis Koutroumanidis, Dr Sushma Goyal and Mr Stuart Smith.

Chapter 6

The SGDP Biobank, and various members of the Pal Lab including myself performed DNA extraction. Exome sequencing was performed and annotated by the Guy's Genomics Facility by their pipeline as described. Amie Jaye supported me in assembling the annotation script for annotation analysis. Simon Topp supported splice site annotation. Case control association analysis using EPACTS script was originally written by Professor Michael Simpson. Savannah Ivy, research assistant at the Pal Lab, performed some PCR reactions. Sanger sequencing was performed at other centres where stated. All remaining work is my own.

Chapter 7

This chapter is my own work.

Chapter 8

This chapter is my own work.

Abbreviations

Abbreviation	Meaning
3di	The Developmental, Dimensional and Diagnostic Interview
1000G	1000 Genomes Project
ABAS	Adaptive Behaviour Assessment System
ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
bp	Base pairs
CADD	Combined Annotations Dependent Depletion
CBRS	Conner's Behavioural Rating Scale
CNV	Copy number variant
DNA	Deoxyribonucleic acid
EE	Epileptic encephalopathy
EEG	Electroencephalogram
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
GGE	Genetic generalised epilepsy
gnomAD	genome Aggregation Database
GSW	Generalised spike wave
GTCS	Generalised tonic clonic seizure
ID	Intellectual disability
IGV	Integrative Genomics Viewer
ILAE	International League Against Epilepsy
IQ	Intelligence quotient
JME	Juvenile Myoclonic epilepsy
LGS	Lennox Gastaut syndrome
MAE	Myoclonic astatic epilepsy
MAF	Minor Allele Frequency
NCBRS	Nicolaides-Baraitser syndrome
NGS	Next generation sequencing
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PP2hvar	Polymorphism Phenotyping v2 HumVar trained
PPR	Photoparoxysmal reactions
RVIS	Residual Variation Intolerance Score
RR	Relative risk
SCQ	Social Communication Questionnaire
SD	Standard deviation
SDQ	Strength and Difficulties Questionnaire.
SIFT	Sorts intolerant from tolerant
SMEI	Severe myoclonic epilepsy of infancy
VUS	Variant of uncertain significance
W_M_XXXXXXX	WTSI_MAE_XXXXXXX

Chapter 1 Introduction: The phenotype and genetic basis of myoclonic astatic epilepsy

Myoclonic astatic epilepsy (MAE) is a rare childhood epilepsy syndrome regarded as having a genetic aetiology. MAE was first described by Herman Doose in 1970 as a category of epilepsy characterised by onset of primary generalised seizures in the form of myoclonic and astatic seizures, often combined with absences, tonic clonic and tonic seizures[1]. However, the phenotypic manifestations and nosological boundaries of MAE are frequently debated[2-4]. The clinical continuum of MAE overlaps with other epilepsy syndromes like benign myoclonic epilepsy of infancy or Lennox Gastaut syndrome (LGS), and the precise determination of myoclonic or atonic falls essential in forming a diagnosis, is difficult without combined electroencephalography and electromyography recordings.

The importance of genetic factors in MAE was recognised with its first description through the evidence offered from family history and family electroencephalogram (EEG) studies[5]. Early twins studies[6,7] and rare mendelian pedigrees[8-10] offered a glimpse to the genetic premise of MAE but the genetic aetiology remained largely elusive till the advent of next generation sequencing (NGS). The quest for gene discovery had lead to 10 genes; *SCN1A*[10,11], *GABRG2*[9,12], *SLC2A1*[13], *CHD2*[12,14,15], *SYNGAP1*[12,16], *SLC6A1*[17,18], *KCNA2*[19], *STX1B*[20,21], *TBC1D24*[22] and *KIAA2022*[23] currently associated with MAE. However, cumulatively these genes account for less than 20% of cases leaving the majority still unexplained.

In this chapter, I shall first trace the historical evolution of concepts around MAE, describe the seizure phenotype of MAE and explore the associated neurodevelopmental comorbidity. Next, I shall review the evidence for genetic influences; through family history and family EEG studies, rare mendelian pedigrees, identification of causative monogenic genes and structural genomic variation. Last, I shall propose strategies that may be helpful in dissecting out major components in its genetic aetiology.

1.1 Historical perspective

Description and refinement of MAE

Before the description of MAE, all such cases were probably categorised as LGS, a syndrome that was originated by Lennox and Davis in 1950[24]. These authors described the slow spike and wave EEG pattern and correlated it with clinical manifestations of myoclonic jerks, atypical

absences, astatic seizures and mental retardation. Subsequently in 1966, Henri Gastaut described the "Lennox syndrome or childhood epileptic encephalopathy with diffuse slow spike and waves" in 100 patients with diffuse slow spike wave on EEG, frequent tonic seizures, absences, sometimes myoclonic seizures and mental retardation[25]. The contributions of both groups were recognised and the term LGS coined at the American Electroencephalographic Society proceedings in 1969.

The concept of separating epilepsy with myoclonic seizures emerged around the same time when Harper described 14 children with myoclonic epilepsy distinct from LGS[26]. Kruse described a kind of epilepsy characterised by myoclonic and astatic seizures under the heading of 'myoclonic astatic petit mal' among other petit mal epilepsies[27]. Doose differentiated this further and in 1970 published 51 cases of "centrencephalic myoclonic astatic petit mal" characterised by onset of primary generalised seizures in the form of myoclonic and astatic seizures, often combined with absences, tonic-clonic and tonic seizures in children between 1 and 5 years old. The prominent seizure symptom was an abrupt loss of postural tone. Doose photographically captured this abortive astatic seizure shown in Figure 1.1, the child's head drops forwards and the knees sag, there is myoclonic flexion of the arms and loss of postural tone. The EEG usually showed bilateral synchronous spike and wave activity with abnormal background theta rhythm. He recognized that the course of the disease is generally unfavourable and in most instances mental retardation developed[5]. Although attributed to Herman Doose, he never originally described MAE as an epilepsy syndrome per se, but instead as a paradigmatic example of a multifactorial determined disease with uniting features[28].

Figure 1.1. The abortive astatic seizure in Doose *et al.*

The child's head drops forwards and the knees sag, there is myoclonic flexion of the arms at the start of the seizure followed by loss of postural tone.

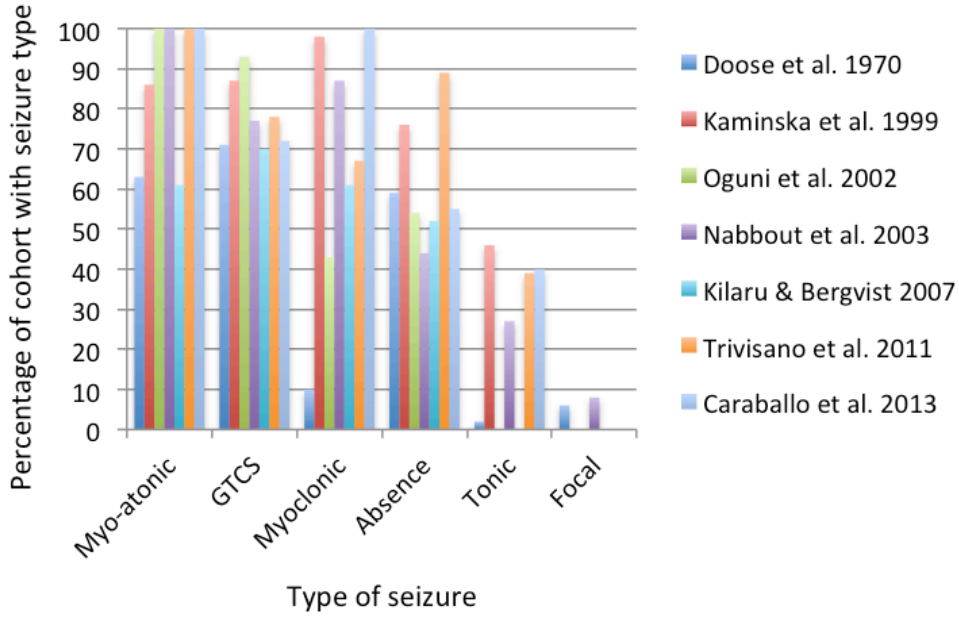
Within this original group of patients reported by Doose *et al*., discrete epilepsy syndromes can now be recognized[5]. Five of his 51 cases had isolated myoclonic seizures that might now fit the definition of benign myoclonic epilepsy of infancy. Ten cases had a seizure onset before the age of 1 year, and 11 cases had febrile convulsions. Although specific details were not differentiated in each patient, these cases could fit a severe myoclonic epilepsy of infancy (SMEI) phenotype. Another subgroup of six cases had evidence of cerebral damage with neurologic symptoms or mental retardation, and one case exhibited frequent tonic seizures; these might better fit the label of LGS. Twenty-two years later, Doose refined his criteria for MAE, adding that tonic seizures were an uncommon feature, and acknowledging the overlap with other epilepsy syndromes[28]. In 1989, the International League Against Epilepsy (ILAE)[29] recognized MAE to have the following features: (1) usually normal development before onset of epilepsy; (2) onset of myoclonic, myoclonic-astatic, or astatic seizures between 7 months and 6 years of age; and (3) presence of generalized spike or polyspike wave EEG discharges. The ILAE also recognized a "hereditary predisposition" with a variable outcome. These criteria are largely based on Doose's original description on a cohort that is now recognized to be phenotypically heterogenous. Nevertheless, seven MAE clinical cohorts have been published and Table 1.1 and Figure 1.2 summarises the main clinical features in these reported cases[2,5,30-34].

Table 1.1. Summary of main clinical characteristics of MAE series.

Cohort	No. of cases	Boys (%)	Febrile seizures (%)	Family history (Ep, FS)%	Mean age of onset (months)
Doose *et al*.1970[5]	51	71	22	40, -	36-48
Kaminska *et al*. 1999[2]	55	78	16	12, -	35.6
Oguni *et al*. 2002[30]	81	75	-	14, 18	32
Nabbout *et al*. 2003[31]	22	-	-	13,18	40
Kilaru & Bergqvist 2007[32]	23	83	17	39, -	36
Trivisano *et al*. 2011[33]	18	89	-	-	43
Caraballo *et al*. 2013[34]	69	67	13	39,11	39

Ep epilepsy, FS febrile seizures, - Not available,

Figure 1.2. Summary of the main seizure types reported in MAE series.



Myo Atonic, consist of myoclonic atonic or atonic seizures, and drop attacks. GTCS generalised tonic clonic seizures, Absence comprise typical and atypical absences

Classification

The concept of an epilepsy syndrome has evolved with the recent ILAE classification[35]. The 1989 ILAE report adopted a broad understanding of the term "syndrome" as an epileptic disorder characterized by a cluster of signs and symptoms customarily occurring together and placed MAE within the category of generalized cryptogenic or symptomatic epilepsies[29]. The 2010 ILAE classification specified that an electroclinical syndrome is a complex of clinical features, signs, and symptoms that together define a distinctive, recognizable clinical disorder. The 2010 report avoids the cryptogenic or symptomatic etiologic distinction by placing MAE (now termed epilepsy with "myoclonic-atonic" seizures, instead of the previously called "myoclonic-astatic" seizures) as a distinct electroclinical syndrome[35]. Although the classification has changed, the diagnostic definitions have not altered with the new classification. Most reports about MAE in the literature use the 1989 ILAE definition. I will from this point on refer to the 1989 ILAE definitions of MAE, unless otherwise specified[29].

1.2 Phenotypic features

1.2.1 Clinical epidemiology

The incidence of MAE is generally low. Amongst newly diagnosed adult and childhood epilepsies over 14 months in a French centre, MAE account for 0.3% of 1016 patients[36]. In paediatric centres, MAE accounted for 1-2% of childhood epilepsies up to the age of 9 years in the German city of Kiel[37], were diagnosed in 18 (5.5%) out of 327 children with a diagnosis of generalised epilepsy in Rome, Italy[33], and 81 (2.2%) out of 3600 children with epilepsy in a Japanese centre[30].

The age of seizure onset ranges from 7 months to 6 years, and peaks between 3 and 4 years. Boys are affected almost twice as frequently as girls (mean of 77.1% are boys in six series)[2,4,5,30,32,34], although when onset is less than one year, there is an equal sex ratio[5]. Doose *et al*. stated that MAE usually begins in previously normally developing children. However, in his original cohort, 6/51 cases had evidence of cerebral damage or delayed development prior to seizure onset[5]. The term 'usually' prior normal development was retained by the 1989 ILAE classification of MAE[29]. The extent of developmental delay in infancy can be difficult to assess as this is largely dependent on parental reportage following presentation of the child with seizures. Still, this distinction is important, as patients with epileptic encephalopathy (EE) and LGS tend to have a developmental delay prior to seizure onset.

1.2.2 Pathology

Generalised subcortical atrophy on cranial computerized tomography has been reported in groups of patients with symptoms similar to MAE prior to the 1989 classification[25]. However, the pathological significance of generalised subcortical atrophy is difficult to determine as it may result form repeated seizures, episodes of status epilepticus, or from hormonal treatment. Subsequent MAE series have indicated no evidence of brain lesions on magnetic resonance imaging[2,4,32]. There are no published post-mortem brain studies in MAE patients.

A EEG-fMRI study offered interesting insights into the pathogenesis of seizure generation in MAE[38]. The authors captured generalised spike wave (GSW) associated BOLD (blood oxygenation level dependent) signal changes in 11 MAE cases. Like other genetic generalised epilepsies (GGE)[39], GSW-associated BOLD signal in MAE cases increased in the thalamus (11/11 cases), premotor cortex (6/11 cases) and putamen (6/11 cases). Deactivation was found in the default mode areas in 7/11 cases. The default mode network consists of precuneus,

retrosplenial cortex, parietal, and anterior medial frontal cortex. It is active in the resting brain with a high degree of functional connectivity; disruption of which is suggested to give rise to epileptic spikes. Additionally GSW-associated BOLD signal activation was seen in brain structures associated with motor function (premotor cortex and supplementary motor area), this focal correlate could be linked to the generation of myoclonic/atonic seizures witnessed in MAE cases.

1.2.3 Seizures

MAE is a generalised epilepsy syndrome and hence patients predominantly present with generalised seizures. Generalised seizures are seizures that originate at some point within and rapidly engage to bilaterally distributed networks[35,40]. A typical MAE seizure history would be of a 3 to 4 year old boy presenting with recurrent febrile or afebrile generalised tonic clonic seizure (GTCS) followed by myoclonic-atonic seizures and variably other generalised seizure types e.g. myoclonic, absences and on going GTCS. Regardless, many children deviate from this presentation. A description of the characteristics of the main seizure types is summarised below.

Myoclonic atonic or atonic seizures

An atonic seizure is defined as a sudden loss or diminution of muscle tone without apparent preceding myoclonic or tonic event lasting ~ 1 to 2 seconds, involving head, trunk, jaw or limb musculature[40]. Whereas a myoclonic atonic seizure is defined as a generalised seizure type with a myoclonic jerk leading to an atonic drop, this was previously called a myoclonic astatic seizure[40]. Myoclonic atonic seizures are the hallmark seizure in MAE and an essential component in its phenotypic manifestation[5,28,30,32]. However, it is difficult to determine whether all reported MAE patients have myoclonic atonic or atonic seizures because of the following: (1) some series group together myoclonic seizures, myoclonic atonic seizures and atonic seizures; (2) researchers use different criteria for myoclonic atonic seizures and (3) it is difficult to qualify the exact physiologic mechanism of drop attacks without combined electroencephalogram/electromyogram recordings. For example Oguni *et al*. reported in his earlier series a cohort with 100% occurrence of either myoclonic or myoclonic atonic seizures[4], subsequently he specified that the intensity of myoclonia and atonia should be equal in order for the term myoclonic atonic seizures to be used; and his later series reported only 10% of cases with myoclonic atonic seizures[30]. Kaminska *et al*. reported drop attacks in 89% of their cohort but did not distinguish this further[2]. Interestingly, precise seizure delineation with EEG analysis of 30 MAE cases into drop attacks caused by myoclonic or atonic groups concluded

that there was no significant impact on clinical outcome[4]. Selection bias may have had a role as the cases were investigated due to refractory seizures requiring multiple hospital admissions.

Myoclonic seizures

A myoclonic seizure is defined as a sudden, brief (<100ms) involuntary single or multiple contraction(s) of muscle(s) or muscle groups of variable topography (axial, proximal limb, distal)[40]. The frequency of myoclonic seizures varies from approximately 43-100% in MAE[2,30,34]. Myoclonic jerks tend to involve the trunk and proximal muscles, can be both flexor and extensor and may cause drop attacks. This corresponds to a GSW with a median frequency of 1.3 Hz on EEG[41]. Aicardi described some patients with the so called myoclonic variant of LGS to have an unusually marked myoclonic component[42]. However a neurophysiologic study demonstrated that myoclonus originates differently in a small cohort of LGS and MAE cases[43]. In three LGS cases, topographic voltage mapping of the pre-myoclonic EEG spike peak showed a lateralised frontal cortex distribution, whereas in three MAE cases, this mapping showed a diffuse distribution of the electrical field, predominantly over the anterior regions but not lateralised, suggesting that myoclonic jerks in MAE is a primary generalised epileptic phenomenon. The occurrence of myoclonic seizures without an atonic component should prompt the consideration of alternate myoclonic epilepsies such as benign myoclonic epilepsy of infancy if onset is less than one year, or severe myoclonic epilepsy of infancy if there is a history of fever sensitivity.

Generalised tonic clonic seizures

A generalised tonic clonic seizure (GTCS) is defined as a bilateral symmetric tonic contraction and then bilateral clonic contraction of somatic muscles, usually associated with autonomic phenomena and loss of awareness[35,40]. These seizures engage networks in both hemispheres at the start of the seizure. Febrile and afebrile GTCS are the first seizure type in more than two thirds of MAE cases[5,30,32]. After myoclonic-atonic or myoclonic seizures, they are the next most frequent seizure reported in the clinical course of MAE[28,32], and more common than absence seizures.

Absences

A typical absence is defined as a sudden onset interruption of on-going activity, with a blank stare and possible brief upward deviation of the eyes, which last a few seconds to a minute with rapid recovery. EEG shows GSW during the event[40]. An atypical absences is an absence seizure with more pronounced changes in tone that a typical absence, the onset and/or cessation of the seizure is not abrupt. EEG often shows slow, irregular GSW activity. Both

typical absence and atypical absence seizures are commonly seen and a whole spectrum of clinical manifestations ranging from typical absences to loss of muscle tone, eyelid myoclonia and sialorrhea is reported[31].

Tonic seizures

A tonic seizure is a sustained increase in muscle contraction lasting a few seconds to minutes[40]. There is no agreement on the frequency of tonic seizures in MAE and figures between 0% and 55% have been reported[2,30,32]. Doose stated that tonic seizures can occur in MAE during sleep, and only in rare cases during daytime[28]. In contrast 75-90% of LGS patients undergoing sleep EEG recording exhibit tonic seizures[44]. Tonic seizures may appear later in the course of LGS rather than at onset, and thus it may be necessary to re-evaluate MAE patients for this symptom to differentiate the two conditions.

Status epilepticus

Status epilepticus was re-defined by the ILAE in 2015[45]. The proposed new definition was as follows: status epilepticus is a condition resulting either from failure of the mechanisms responsible for seizure termination or from the initiation of mechanisms, which lead to abnormally, prolonged seizures (~ after 5 minutes). It is a condition, which can have long-term consequences (~ after 30 minutes), including neuronal death, neural injury and alteration of neuronal networks. In MAE, a 'petit mal status' or 'status of minor seizures' consisting of a series of myoclonic atonic seizures or myoclonus and atypical absences is typical and much more common than convulsive status[5] (ILAE axis B.2.a.b. or B.2.a.c.[45]). This contrasts with LGS in which status characteristically involves clouding of consciousness with frequent tonic seizures (ILAE axis A.1.a. or A.4[45]). Kaminska *et al*. reported myoclonic status as a poor prognostic factor occurring in 14% of favourable MAE and 94.5% of unfavourable MAE[2]

Focal seizures

Focal seizures are defined as seizures that originate within networks limited to one hemisphere. They may be discretely localised or more widely distributed. Focal seizures can be classified into seizure with or without awareness and then along with further manifestation in motor (tonic, clonic, myoclonic, hyper motor etc.) or non motor (somatosensory, olfactory, visual, auditory etc.) categories[35]. Classifications aside, focal seizures are not a frequent seizure type in MAE. Nabbout *et al*. reported 3/22 cases with focal seizures, type unspecified whereas Doose *et al.* reported 3/51 cases with focal seizures (two cases were described as psychomotor and one case had focal motor seizures)[5,31]. Accordingly, focal activity is also uncommon in the EEG.

1.3 EEG

The EEG can show background changes with abnormal centroparietal theta rhythms or can be normal at the start of the epilepsy[5]. Hyper synchronous theta activity was recorded at epilepsy onset in 16/18 cases in Trivisano *et al*.'s cohort[33]. With progression of the disease, bursts of predominantly 2 to 5Hz generalised irregular spike and wave and/or polyspike and wave become prominent[4]. A useful additional marker is the electromyography recording during atonic seizures which demonstrate a slow wave component of the discharge[46]. Although it is recognised that MAE can present as an EE[47], generally posterior background rhythms and sleep architecture are normal, which is in contrast to LGS where there is little or no normal background activity. Additionally although slow spike waves can be seen in both MAE and LGS, in LGS this is often combined with focal abnormalities; whereas in MAE focal activity is unusual and acted as indicators of poor prognosis in a cohort of nine cases[48]. During remission, it is typical for a marked diffuse abnormal theta rhythm to develop[49]. Therefore, although there are no pathognomonic EEG signatures for MAE, sufficient EEG features exist to distinguish from other epilepsies, when taken in conjunction with a consistent clinical history.

1.4 Neurological examination

Neurological examinations of MAE cases have been recorded in three series with little agreement. Doose *et al.* reported 6/51 of his cohort with variable abnormal neurology of cerebral palsy in two, abnormal reflexes in three and facial paresis in one[5]. Nabbout *et al*. reported 3/22 cases with ataxia whereas Caraballo *et al.* determined that all 69 of their cases had normal physical examinations and were not dysmorphic[31,34].

1.5 Neurodevelopmental comorbidity

Remarkably little detail has been reported about the cognitive or behavioural phenotype in patients with MAE and comorbidity is mainly inferred from case series and case reports.

Behavioural comorbidity

Behavioural comorbidities including autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) are common in children with epilepsy[50]. Escayg *et al*. reported hyperactivity and behavioural disturbances in 10/22 MAE patients, although specific measures and difficulties were not specified[10]. Nolte *et al*. followed 15 MAE cases for 8.5 years and reported persistent behavioural abnormalities in five cases; hyperkinesia, short attention span and emotional liability were the most frequent complaints[51]. A few case reports have reported

cases with behavioural difficulties[21,52], such as a child with behavioural inhibition and shyness on the child behavioural checklist[53]. The child behavioural checklist was also employed by Trivisano *et al*. who reported normal scores in 17/18 cases. However, when scores were compared with controls, there was evidence of significant subclinical difficulties in subtest involving symptoms of withdrawn/depression ($P<0.05$), attention behaviour ($P<0.01$) and aggressive behaviour ($P<0.01$). The cause of behavioural difficulties are often multifactorial and can be related to active seizure periods, anticonvulsant medication, or as a manifestation of cognitive difficulties.

Cognitive comorbidity

Cognition can be defined as the capacity of the brain to process information accurately and to programme adaptive behaviour. It involves the ability to solve problems, to memorise information and to focus attention[54]. As a group, 355 individuals with generalised epilepsy had impaired cognitive performance in comparison to 75 healthy subjects matched for age and education ($P<0.0003$)[55]. Children with symptomatic generalised epilepsies (like MAE)[29], frequent seizures, high antiepileptic drug use and early onset seizures appear most at risk[55,56]. In MAE, cognitive prognosis tends to be favourable if seizures remit and unfavourable when seizure persist[57]. Table 1.2 summarise the cognitive profile of MAE cases based on data from published series.

Table 1.2 Summary of cognitive studies in MAE cases

Study	No. of cases	Assessment tool	IQ scores
Doose *et al*. 1970[5]	51	Not specified	34 (66%) cases normal 10 (20%) cases mild 7 (14%)cases moderate
Kaminska *et al*. 1999[2]	55	IQ test or educational attainment	22 (40%) cases >80 17 (30%) cases 50-80 16 (29%) cases <50
Oguni *et al*. 2002[30]	81	Modified Binet IQ or WISC R	48 (59%) cases >80 16 (20%) cases 60-79 17 (21%) cases <59
Kilaru & Bergqvist 2007[32]	21	Normal – mainstream education Mild – mainstream with help Moderate – special education	9 (43%) cases normal 11 (52%) cases mild 1 (5%) case moderate
Trivisano *et al.* 2011[33]	18	Leiter International Performance Scale Revised	11 (61%) cases >80 5 (27%) cases 60-79 2 (11%) cases <59
Normal or IQ >80: mean 53.8% SD 11.5, median 59% range 40 – 66% Mild or IQ 60* to 79: mean 29.8% SD 13.1, median 27% range 20 - 52% Moderate of IQ <59*: mean 16% SD 9.2, median 14% range 5 – 29% IQ scores divided as authors categories[2]			

Important methodological issues should be mentioned with these cognitive results; (1) cognitive testing at any time point can be influenced by current seizure burden and antiepileptic drug use, (2) it is difficult to measure the cumulative effect of epilepsy duration prior to cognitive testing, (3) there is no consistent data on pre-seizure cognitive levels and problems may have already been present, (4) different investigators use different psychometric tools, (5) there is no account on the influence of psychosocial factors which may play an important role and (6) studies are mostly limited to measures of intelligence (IQ). Nevertheless, a spectrum of cognitive abilities is seen with 40 – 66% of cases classified as normal or IQ >80. Trivisano *et al*. correlated cognitive delay with family history of epilepsy, tonic seizures, epileptic encephalopathy, epilepsy duration, age of onset (P=0.06) and non convulsive status epilepticus (P=0.07), and found no statistical differences[33]. Other investigators have attempted to divide MAE into favourable and unfavourable groups.

1.6 Course and Prognosis

Prognosis is variable in MAE. Doose identified the following risk factors for an unfavourable prognosis: onset with febrile and afebrile GTCS during the first 18 months of life, status of minor seizures, persistence of 4 to 7Hz and failure to develop a stable occipital alpha rhythm[28]. Three further studies have attempted to differentiate MAE into favourable and unfavourable groups; favourable cases had seizure remission with no evidence of cognitive impairment[2,30,34]. Using a method of data reduction known as multiple correspondence analysis, Kaminska *et al*, found that both groups were indistinguishable at onset but those with poor outcome after 3 years demonstrated lack of familial antecedents, tonic and absence seizures, myoclonic status and long bursts of irregular spike and slow wave on EEG[2]. Oguni *et al*. retrospectively compared 55 favourable and 15 unfavourable MAE cases and reported that a positive family history of epilepsy and absence status or minor epileptic status significantly correlated with an unfavourable outcome[30]. Caraballo *et al*, was able to distinguish his cohort into 39 favourable and 30 unfavourable cases based on electroclinical features and evolution. The key distinguishing features in the unfavourable group which had occurred within 10 month of epilepsy onset were; myoclonic and absences status, tonic seizures and frequent generalised spike and polyspike wave discharges and focal spikes on EEG[34].

Putting together the evidence, cases with unfavourable outcome might be characterised by absence or minor epileptic status, and tonic seizures – all features that shift the overall picture toward LGS. However when features in groups of patients with MAE and LGS were compared, there was no transition in syndrome diagnosis between the two groups, even though EEG

patterns were changeable and some electroclinical features overlapped[58]. These studies indicate that there are identifiable but not completely reproducible risk factors for an unfavourable prognosis that may not be present at onset.

Seizure remission

In retrospective studies, seizure remission between 12 to 24 months was achieved in two thirds of MAE patients[2,30,34]. Myoclonic and/or atonic seizures disappeared in 89% within one to three years but GTCS tended to continue in one cohort[30]. It is possible that seizure remission rates may be underestimated because of selection bias in follow up studies.

1.7 Genetic Aetiology

Evidence supporting a genetic cause of MAE comprises twin studies, family studies (some with EEG), and gene mutation analyses. Twin studies generally compare monozygotic and dizygotic twins through a comparison of disease concordance rates. This design is based on the fact that monozygotic pairs are assumed to be genetically identical in DNA sequences, whereas dizygotic twins share approximately 50% of their gene sequence. Genetically influenced characteristics may show a higher concordance in monozygotic than dizygotic twins, assuming that both types of twins equally share environmental influences. Family studies examine the distribution of traits among members of a family, but cannot distinguish shared genetic and environmental factors when familial aggregation is present. Genetic linkage studies are used to identify regions of the genome that contain genes that predispose to disease by observation of allele sharing among related individuals. Gene mutation analyses correlate mutations in genes with particular phenotypes comparing cases with controls and sometimes with relatives.

1.7.1 Twin studies

Twin studies of symptomatic or cryptogenic generalised epilepsies[29] have shown a higher concordance rate of 83 to 94% in monozygotic pairs compared to 65 to 71% in dizygotic pairs[6,7]. A discordant monozygotic pair was reported within one of these studies, in which one twin had MAE and his co-twin had an unclassified epilepsy[6]. The presence of this monozygotic twin pair concordant for epilepsy and having different epilepsy syndromes might be explained by epigenetic or environmental factors interacting with a genetic susceptibility. Family studies also suggest that the underlying genetic susceptibility may not be to the syndrome itself.

1.7.2 Family studies – seizures

A positive family history of seizures in first, second or third degree relatives is seen in about one third of cases[5,32,34]. In Doose's series, the incidence of afebrile or febrile seizures was higher in siblings (brothers 18%, sisters 12%) than in parents (father 7%, mother 4%), although these differences were not statistically significant (parents' *P*=0.36; siblings' *P*=0.38)[1] (see Table 1.3). This unusual distribution is seen in some idiopathic generalised epilepsies with genetic anticipation[59], but is more likely to be due to recall bias for seizures in the parental generation[60]. The seizure types present in the relatives were predominantly febrile or afebrile GTCS and myoclonic or myoclonic atonic seizures occurred in only three of 160 siblings[5]. The low familial aggregation of myoclonic or myoclonic atonic seizures is an interesting observation; one hypothesis is an interaction between an inherited susceptibility to generalised seizures with, a separate and low frequency (genetic, epigenetic, or environmental) factor modifying the phenotype[61]. This situation is seen in the differential inheritance of different seizures types in idiopathic generalised epilepsy families[62,63]. Familial EEG studies offer further clues about the exact nature of inherited traits in MAE families.

Table 1.3. Familial prevalence of seizure history in families of 107 cases with MAE.

Class of relative	Number of individuals (%)
Brothers	78 (18)
Sisters	82 (12)
Father	100 (7)
Father's siblings	246 (3)
Mother	102 (4)
Mother's siblings	221 (4)

1.7.3 Family studies – EEG

Herman Doose was aware of the heritable nature of MAE and carried out EEG family studies in 72 siblings and 79 parents of MAE probands with his first description of MAE in 1970[5]. Doose defined specific EEG abnormalities evaluated as (1) photoparoxysmal responses (PPR), with bilateral synchronous or only parietal occipital irregular spike and waves during photic stimulation, (2) GSW at rest or during hyperventilation and (3) dysrhythmia defined as an irregular background activity with widely varying amplitudes and frequencies such that no dominant activity is recognisable. He went on to detail that in dysrhythmia, the amplitude of the EEG may vary considerably; sharp transients may be interspersed but not spikes and distinct sharp waves. Dysrhythmia may be continuous or may occur in paroxysmal groups. Table 1.4 summaries the findings from these EEG family studies. The most common EEG

abnormality was PPR and there seemed to be a clear female excess; present in 16 of 39 sisters compared to 4 of 33 brothers (*P*<0.0025). There was also evidence of preferential maternal transmission of the EEG traits: 9 of 37 mothers as opposed to 1 of 32 fathers of MAE probands had a "pathological" EEG.

Table 1.4 EEG findings in MAE familes performed by Doose *et al*.

EEG findings	Siblings (n=72)	Fathers (n=32)	Mothers (n=37)	Total
PPR	20 (27.8%)	-	6 (16.2%)	26 (18.5%)
GSW	6 (8.3%)	1 (3.1%)	-	7 (4.9%)
Dysrhythmia	14 (19.4%)	-	3 (8.1%)	17 (12.0%)
Total	33 (45.8%)	1 (3.1%)	9 (24.3%)	50 (35.4%)

PPR photoparoxysmal responses, GSW generalised spike wave

The maternal preponderance of EEG abnormalities, along with the male bias for MAE in probands is intriguing and has several possible genetic explanations (see Table 1.5)[63].

Table 1.5. Patterns of maternal transmission and expected offspring sex ratio.

Maternal transmission	Sex ratio in offspring
X linked recessive	Only boys affected
X linked dominant[a]	Equal
Mitochondrial	Equal with variable expression
Triplet repeat[a]	Equal with anticipation
Ascertainment bias	Female or male excess
Non-paternity	Equal
Perinatal or pregnancy factors	Variable depending on factors
Paternal imprinting	Equal
Sex dependent penetrance	Female or male excess

[a]Paternal transmission possible

Abnormal EEG findings occurred mostly in siblings (45.8%) and were significantly higher when compared to 8.3% of kindergarten and school aged controls (*P*<0.01). Additionally in siblings, the younger age groups (<10 years) had more frequent abnormalities particularly in dysrhythmia and PPR, possibly capturing age related EEG features (see Table 1.6). Moreover, GSW, the prototype epileptiform discharge were the least reported group overall (4.9%). However, It is also worth noting that 34 (68%) families were recorded as having a history of convulsions in first or second degree family members and five siblings, one father and two mothers had a history of seizures.

Table 1.6 EEG findings in MAE siblings by Doose *et al*.

EEG findings	1 to 5 years (n=21)	6 to 10 years (n=25)	11 to 15 years (n=16)	15 years (n=10)
PPR	3 (14.2%)	11 (44.0%)	3 (18.7%)	3 (30.0%)
GSW	3 (14.2%)	1 (4.0%)	2 (12.5%)	-
Dysrhythmia	9 (42.9%)	4 (16.0%)	-	1 (10.0%)
Total	11 (52.4%)	14 (56.0%)	4 (25.0%)	4 (40.0%)

PPR photoparoxysmal responses, GSW generalised spike wave

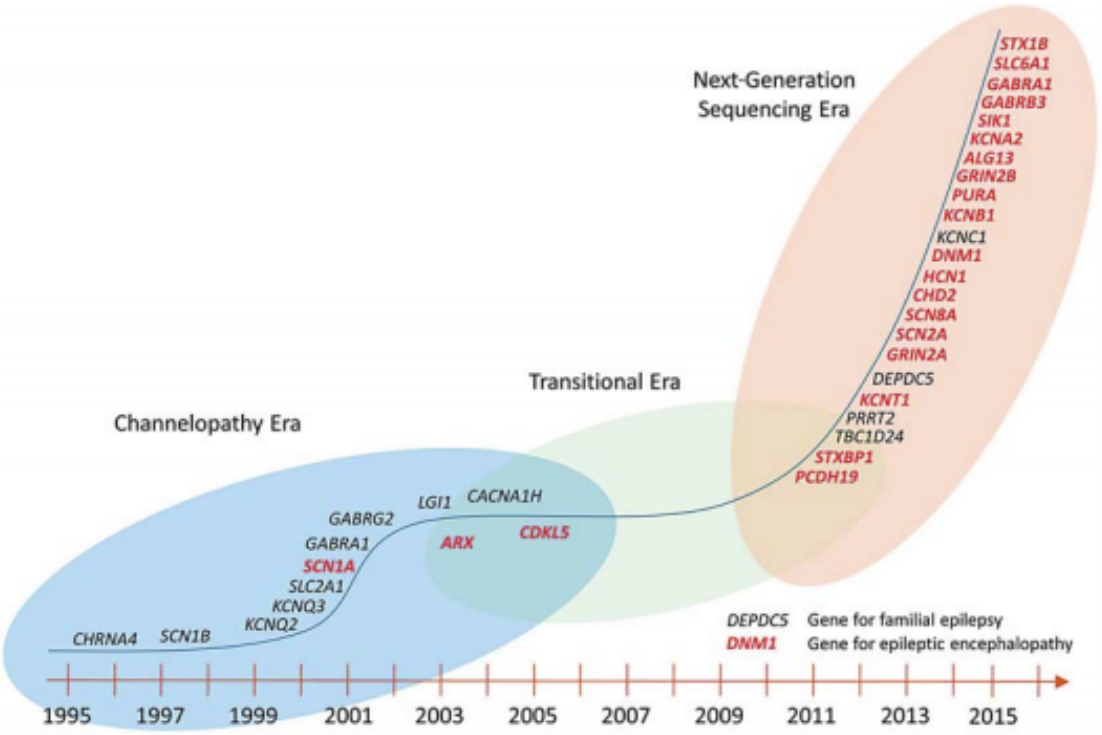
Doose *et al*. showed that of the 50 families, taking both seizure and EEG data into account 34 (68%) of families had members (excluding the proband) with evidence of seizure susceptibility[5]. These 34 families included 11 families with both pathologic EEG and seizure history, 12 families with pathologic EEG only, and 11 with seizure history only. As not all ascertained families had EEG studies, these cases may have been subject to selection bias.

EEG features may provide a useful endophenotype in family based genetic studies to shed light on the genetic model. Criteria used for an endophenotype in psychiatric genetics include the following: (1) the endophenotype is associated with illness in the population, (2) the endophenotype is heritable, (3) the endophenotype is state independent (manifest in an individual whether or not illness is active), (4) the endophenotype and illness co-segregate within families, and (5) the endophenotype found in affected family members is found in non affected family members at a higher rate than in the general population[64]. PPR and GSW easily satisfy the criteria for endophenotypes and have been employed before[65-67], but theta rhythms may prove more problematic because of their resemblance to physiological EEG phenomena[68].

1.7.4 Gene mutation studies

This section has been divided into the findings pre next generation sequencing (NGS) and post NGS. This distinction is important, as the NGS era has brought about an explosion of new gene associations. Figure 1.3 by Helbig *et al.* shows a visual representation of this phenomenon[69].

Figure 1.3 Timeline of gene discovery in human epilepsies from Helbig *et al*.



1.7.4.1 Pre Next Generation Sequencing

Rare Mendelian pedigrees

Prior to NGS, the success of identification of disease-linked variants has been largely limited to large well-characterised pedigrees. Genetic studies in phenotypically heterogenous multiplex epilepsy families have revealed several different mutations in MAE patients. *SCN1A*[10,70] and *SCN1B*[8] mutations segregated in three separate generalised epilepsy with febrile seizures plus (GEFS+) pedigrees that included one family member with a MAE phenotype. Additionally, linkage analysis identified *GABRG2* in a large epilepsy pedigree with childhood absence epilepsy and febrile seizures[9]; the pedigree has a single member with MAE although his genotype was not investigated. In each of these pedigrees, only one member was affected with MAE, whereas other members had diverse seizure phenotypes. This suggests that the association of MAE with *SCN1A*, *SCN1B* and *GABRG2* may be purely private to these pedigrees.

Sporadic cases

This hypothesis of a private mutation was tested in 22 sporadic French MAE cases and four Japanese MAE cases. *SCN1A*, *SCN1B* and *GABRG2* variants were not found through standard PCR and Sanger sequencing of exons[11,31]. However, a novel *SCN1A* frameshift variant ending in a premature stop codon L433fsX449 has been reported in a child with overlap features of MAE

and SMEI (unilateral tonic clonic seizures in first year of life, provoked by fever)[71]. Additionally, a recurrent *SCN1A* variant p.T1174S was identified in a boy with sporadic MAE, but the pathogenicity is uncertain as the same variant was identified in his asymptomatic mother[72], and is present in current population databases (gnomAD MAF 1.72 x 10^{-3}). Around the same time, the spectrum and variable expressivity of *SCN1A* related seizures were emerging, and the principally associated phenotypes of SMEI and GEFS+ were recognised[70]. Putting together the evidence, *SCN1A* variants can occur in MAE but are very rare and usually associated in cases with overlap phenotypes of MAE/SMEI or GEFS+.

Outside these initial pedigree associated genes, the next gene identified in MAE cases, *SLC2A1*, offered the possibility of precision medicine. The *SLC2A1* gene encodes the glucose transporter type 1 protein that is responsible for transporting glucose across the blood brain barrier. *SLC2A1* is associated with glucose transporter type 1 deficiency syndrome[73]. The classic phenotype is characterised by infantile onset seizures, delayed neurologic development, acquired microcephaly and complex movement disorders. Treatment with the ketogenic diet is highly effective in improving seizures, movement disorder and alertness. Four our of 84 MAE cases were identified with *SLC2A1* variants using multiplex ligation dependent probe amplification analysis[13]. Three of these four MAE patients had an abnormal neurologic examination with movement disorders. Movement disorder is unusual in MAE and may offer a phenotypic clue to this association. However, with the help of NGS, *SLC2A1* variants were not identified in 120 MAE cases[18] altering the recognised genetic contribution of this gene in MAE, whilst *GABRG2* was reignited as an epilepsy/MAE gene[74]. This and other NGS findings will be discussed next.

1.7.4.2 Next generation sequencing studies

NGS or high throughput/massive parallel sequencing has been a transformative technology and involves the generation of sequence data from hundreds of millions of short DNA fragments in parallel. The first commercial machine for this purpose was available in 2008. This reduced cost and increased speed in generating sequence data lead to a rapid increase of genes connected to MAE. Several different methodologies and commercial machines to generate NGS data are available. The Illumina Genome Analyzer IIx is a widely used method and Table 1.7 compares this with Sanger sequencing.

Table 1.7 Comparison of Sanger sequencing and NGS.

	Sanger sequencing	**NGS (Illumina GAIIx)**
Cost	£0.01 per base pair	£0.0000001 per base pair
Speed	~1.2Mb per day	~0.8Gb per day
Read length	~1000bp	35bp - 400bp

Many of the NGS gene identification studies in MAE have been in the context of investigating EE. The ILAE defined an EE as a condition in which epileptic activity itself contributes to severe cognitive and behavioural impairments, beyond what might be expected from the underlying pathology alone, and that these impairments can worsen over time[35]. This category or process includes a heterogenous group of epilepsy syndromes of early onset with usually intractable seizures, aggressive paroxysmal activity on EEG, and associated with a significant impact on neurological development, cognition and behaviour. When unfavourable, MAE can present as an EE[42,69], and hence have been included within EE cohorts.

All the MAE genes identified from the NGS era, were published during the course of this project, some with overlapping cohorts to this project. Table 1.8 summarises the main features of these genes and the number of published cases.

Table 1.8 Summary of MAE associated genes as of January 2017.

Gene	Gene function	Phenotypic clues	Gene +/ All MAE cases (whole cohort)
SCN1A	Neuronal voltage gated sodium channel	Clinical overlaps with SMEI	1/20[71]
GABRG2	Neuronal $GABA_A$ receptor activity	-	1 case[9] 1/81 (500)[12] 1 case[74]
SLC2A1	Glucose transmembrane transporter activity	Movement disorders	4/84 cases[13]
CHD2	SWI/SNF related matrix associated actin dependent regulator of chromatin	Atonic-myoclonic-absence seizures Clinical photosensitivity with self induced seizures	2/81 cases (6/500)[12] 1/20 cases[14] 1 case[15]
KCNA2	Voltage gated potassium channel	Focal seizures Multifocal epileptiform EEG discharges	1/39 cases (6/255)[19]
STX1B	Helps tethers synaptic vesicles at the presynaptic membrane and mediates neurotransmitter release	Multiple seizure types with febrile and afebrile seizures	2 cases[20] 1 case[21]
SYNGAP1	Glutamate receptor regulation and synaptic transmission	Early developmental delay, progressive ID, delayed speech, autism, hypotonia, unsteady gait	3 cases (17/251)[16]
SLC6A1	GABA transporter responsible for reuptake of GABA from the synapse	Early developmental delay, neurodevelopmental comorbidity	6/160 cases[17] 1 case[75]
TBC1D24	Regulation of synaptic vesicle trafficking	-	1 case (48)[22]
KIAA2022	X linked gene Unknown function	Associated intellectual disability	1 case (14 females)[23]

SMEI severe myoclonic epilepsy of infancy, GABA gamma aminobutric acid, SWI/SNF SWItch/Sucrose Non-Fermentable, ID intellectual disability

CHD2

CHD2 was first introduced as a candidate gene for EE following a screen for *de novo* variants in 264 cases (infantile spasms n=149, LGS n=115) by the Epi4K consortium[76]. Subsequently targeted resequencing of 500 cases with EE identified six cases with *de novo CHD2* variants, of these 2/81 cases had MAE[12]. *CHD2* encodes for chromodomain helicase DNA binding protein 2, it plays an important role in modulating chromatin structure, cell cycle regulation, development and cell differentiation. Pathogenic variants appear to be loss of function as associated cases mostly have copy number deletions or frameshift leading to nonsense mutations[77]. The role of *CHD2* haploinsufficiency was tested in knocked down chd2 zebrafish larvae which exhibited altered locomotor activity and epileptiform discharges, which were absent in control larvae[77].

Most patients with *CHD2* pathogenic variants do not have a MAE phenotype. Indeed the phenotypic spectrum extends to LGS, SMEI, GGE, epilepsy with myoclonic absences, other unclassified EE and ID without epilepsy[15,77-79]. Moreover, *CHD2* variants were evaluated as an independent risk factor for photosensitivity in epilepsy occurring in 11/580 individuals with photosensitive epilepsies compared with 128/34427 controls (P=2.71 x 10^{-5})[78]. Thomas *et al.* delineated the phenotype of nine individuals with childhood EE and *de novo CHD2* variants. He elucidated a phenotypic spectrum of myoclonic epilepsy: marked clinical photosensitivity with self induced seizures with the television, atonic-myoclonic-absence seizures (an abrupt atonic head nod or fall with simultaneous eye deviation progressing to a myoclonic absence phase) and variable ID. The EEG showed GSW and polyspike wave activity[15]. Hence *CHD2* associated epilepsy has overlap features with a MAE phenotype but also distinctive clinical features.

SYNGAP1

SYNGAP1 was identified in a similar fashion to *CHD2*, first through *de novo* screening in 264 cases with EE[76], and then subsequently through targeted resequencing in 500 cases with a spectrum of epilepsy syndromes that can cause EE[12]. The latter cohort identified 5/500 cases with *de novo SYNGAP1* variants but this did not include any of the 81 MAE cases within the cohort. Subsequently Mignot *et al.* identified loss of function *SYNGAP1* variants in 17/251 patients with neurodevelopmental disorders, of these three cases had a MAE phenotype[16]. *SYNGAP1* encodes the synaptic RAS-GTPase activating protein 1, a protein of the post synaptic density of glutamatergic neurons. SYNGAP1 is able to positively or negatively regulate the density of N-Methyl-D-aspartic acid and α-amino-3-hydroxy-5-methyl-4-isoxazoleproprionic acid receptors at the glutamatergic synapses.

Developmental delay is often the first manifestation of *SYNGAP1* associated epilepsy. Atonic, myoclonic and absence seizures are the most common seizure types reported[16]. Progressive intellectual disability (ID) and autism is frequent, as are neurological features of hypotonia and unsteady gait. The clinical phenotype of *SYNGAP1* associated epilepsy in the context of MAE leans towards a severe EE.

KCNA2

KCNA2 encodes the potassium channel Kv1.2, these channels are highly expressed in the central nervous system and have an important role in neuronal excitability and neurotransmitter release. *KCNA2* was first identified as a candidate gene following detection of a *de novo* variant in 33 patients with epilepsy as part of a pilot study[80]. Subsequently several parallel studies investigating 255 individuals identified six further patients with previously unreported lost of function or gain of function *de novo* variants in *KCNA2,* of these were 1/39 MAE case[19]. *KCNA2* associated epilepsy can have a broad phenotypic spectrum from a GGE to an EE[81]. However in Syrbe *et al.*'s cohort, most cases had an early normal development followed by onset of focal and generalised seizures with focal or multifocal sharp waves and spikes on EEG[19]. The only MAE case associated with *KCNA2* had these features plus myoclonic and myoclonic atonic seizures[19].

STX1B

STX1B encodes syntaxin 1b, a component of the SNARE complex, that helps tethers synaptic vesicles at the presynaptic membrane and mediates neurotransmitter release. *STX1B* was first identified in two large German pedigrees with familial febrile seizures. Linkage analysis isolated a single locus on chromosome 16p11.2 with a LOD score of 4.27 leading to the identification of *STX1B* following exome sequencing more than a decade later[20]. Subsequently seven cases with *STX1B* variants have been reported[20,82]. Three of these seven reported cases have a MAE phenotype. The clinical phenotype in MAE cases with *STX1B* consists of early onset seizures, often in the context of a febrile illness with febrile seizures followed by generalized seizures (GTCS, atonic), and with moderate to severe ID.

SLC6A1

SLC6A1 encodes GAT-1, a voltage dependent gamma aminobutyric acid (GABA) transporter that is responsible for the re-uptake of GABA from the synapse. Some have described *SLC6A1* as the "Doose gene" as it has only been linked to MAE patients and at 4% (6/160 cases), it is the most enriched causative MAE gene. Carvill *et al*. published six pathogenic *de novo* heterozygous *SLC6A1* variants, two truncation and four missense[17]. The authors postulated a

loss of function role in the nonsense variants and clustering of the four missense variants around the GABA binding pocket, which in three dimensional space disrupts GABA transport from the extracellular space into the pre-synaptic terminal. Subsequently another MAE case with a *de novo* p.C164T variant was identified, high resolution structural modelling demonstrated disruption of a disulphide bond necessary for stabilization of the transmembrane helix.[75]. Recently, a parental mosaic transmission of a *SLC6A1* p.G263R variant to a child with EE was described, but the specific epilepsy phenotype was not available[83].

SLC6A1 associated MAE tends to be severe, developmental delay is usually present prior to seizure onset, and subsequent ID of varying severity is present. Seizure types are atonic as well as with myoclonic and absence seizures. EEG shows generalised epileptic activity. Psychiatric comorbidity is present in the majority.

TBC1D24

TBC1D24 (TBC1 domain family member 24) is involved in regulation of synaptic vesicle trafficking in brain and somatic development. *TBC1D24* was first mapped in a consanguineous Arab Israeli family with focal epilepsy and ID using multipoint linkage analysis[84]. Since then, several other reports including a recent review on 48 cases has been published; the associated phenotypic spectrum is vast, ranging from isolated deafness, benign myoclonic epilepsies, to EE and severe developmental delay[22]. All reported variants are bialleleic.

A single case with MAE was reported in this cohort[22]. He carried a 16p13.3 duplication of 407kb including but not interrupting *TBC1D24* (2481289-2888632) x3[22]. This case (subject 00523/S2388) was also recruited to this study and exome sequencing did not identify other variants of interest. His phenotype consists of seizure onset at 3 years of absences, atonic, GTCS and myoclonic seizures. He has moderate ID and continues to have drug resistant epilepsy. It is not possible to generalise the *TBC1D24* associated MAE phenotype from this one case.

KIAA2022

KIAA2022 is an X linked ID gene which is highly expressed in the brain, not much is known about its gene function and it is thought to have a role in early brain development. It was first described in hemizygous males[85] and then subsequently in symptomatic female carriers[86]. Lately, Lange *et al*. presented 14 females with *de novo* heterozygous variants. ID was present in all and epilepsy in 12/14, including one patient with MAE[23]. This one case had seizure onset

at 30 months with myoclonic, tonic, atonic and focal seizures. EEG showed polyspike wave and focal discharges. She had ID and ADHD and growth retardation below two standard deviations.

GABRG2

GABRG2 encodes $GABA_A$ receptor γ2-subunit and is part of the family of $GABA_A$ receptor subunit genes (*GABRA1*, *GABRB3, GABRG2* and *GABRD*), all of which have been associated with various epilepsy syndromes. *GABRG2* was first connected to MAE in the pre NGS era from a large epilepsy pedigree with childhood absence epilepsy and febrile seizures using linkage analysis[9]. The pedigree has a single member with MAE although his genotype was not investigated. Subsequently, variants in *GABRG2* has been related to patients with GEFS+ and SMEI[87]. A role as a genetic risk factor has also been explored. Rare variant association analyses demonstrated statistical enrichment of rare *GABRG2* variants in febrile seizures and rolandic epilepsy[88,89].

More recently, *de novo* missense *GABRG2* variants were identified in patients with EE[12,74]. 1/81 MAE case out of 500 cases with EE was identified, this case had seizure onset at 8 months of febrile seizures, myoclonic jerks, atonic, absence and tonic clonic seizures. He was reported to have a normal EEG and a normal cognitive profile[12]. In addition, a review of eight patients with EE and *de novo GABRG2* variants had a single case (patient 6) whose phenotype could fit MAE, although no epilepsy syndrome was specified[74]. These eight cases had a severe phenotype with seizure onset in the first year of life, leading to intractable seizures and severe global developmental delay.

The *GABRG2* associated epilepsy phenotype appears diverse and it is difficult to generalise with only two published MAE cases with quite different profiles. The role of *GABRG2* in MAE is also supported by the occurrence of a *de novo* 5q33-q34 copy number variant (CNV) encompassing *GABRG2* and *GABRA1* in a MAE case[90]. The role of CNVs further complicates the genetic architecture of MAE.

1.7.5 Copy number variants

The role of rare genomic copy number variations (CNVs) in epilepsy is now recognised as a pathogenic principle with specific genomic hotspots (deletions at 15q11.2, 15q13.3 and 16p13.11) predisposing to neuropsychiatric phenotypes and GGE[91,92,93]. In a study of 279 patients with epilepsy and 246 healthy controls, 9.3% of patients with epilepsy had rare CNVs; which were significantly larger (>1Mb, *P*=0.002), had more than 10 genes (*P*=0.005), and were more enriched when associated with mental retardation and neuropsychiatric features (*P*=0.004)[94]. This increased complexity in genetic architecture may add an additional risk factor in disease mechanism. The role of CNVs as a direct causative factor has also been explored.

Four studies have investigated CNVs in MAE cases[90,93,95,96], all used array comparative genomic hybridization in analysis. Table 1.9 summarises the positive findings in these studies. CNVs, which have been reported as part of specific epilepsy gene studies e.g. in CNVs in *TBC1D24* and *STX1B*, have not been included in this table. CNVs were reported rare if they were not present in control populations. Rare CNVs occurred in 9.2% (9/97) and were considered pathogenic in 5.1% (5/97) of MAE cases. The size of CNVs ranged from 85.8kb to 9.45Mb. CNVs were considered pathogenic by investigators if they were recurrent, overlapped with a gene associated with epilepsy and/or were *de novo*.

Table 1.9 Rare CNVs published in MAE patients.

Study (CNV+ MAE/ All MAE cases)	Rare CNV	Size	Overlapping genes	Causal?
Mefford *et al*. 2010[93] (2/15)	5p15.33 dup inh M^	713kb	*NKD2, SLCA18*	L
	7q36.1 del inh P	85.8kb	*GALNT11*	U
Mefford *et al.* 2011[90] (6/77)	15q11 del *de novo*	270kb	*UBE3A*	P
	7q21 del inh M^	3.9Mb	*CACNA2D1, PCLO*	L
	10p13 del *de novo*	5.25Mb	-	P
	5q33-q34 del *de novo*	6.45Mb	*GABRA1, GABRG2*	P
	15q13 del inh P	1.7Mb	-	U
	4q35 dup inh P	1.05Mb	-	U
Helbig *et al*. 2014[95] (1/5)	16p13.11 del *de novo*	-	-	P
Ottaviani *et al.* 2015[96] (case report)	4q21.22-q23 dup *de novo*	778Kb	11 genes	P

Dup duplication, del deletion, inh inherited, M maternal, P paternal, ^symptomatic, L likely pathogenic, U uncertain, P pathogenic, - data not provided, causality based on author's classification.

1.8 Challenges and Strategies

During the course of this project, the number of genes associated with MAE has rapidly increased from two to ten. This is largely due to the success of NGS along with assembly of large cohorts through international collaborations (Epi4K, Euroepinomics) leading to increased power of detection. Although the number of genes have expanded, the numbers of affected cases are still small, uncovering the extent of genetic heterogeneity in MAE.

Collectively, these 10 genes account for approximately 20% of genetic aetiology. This list comprises contribution of pathogenic variants in combined published MAE cases in *SCN1A* 2.1% (1/46), *GABRG2* 0.93% (1/107), *SLC2A1* 1.9% (4/204), *CHD2* 2.9% (3/101), *KCNA2* 2.5% (1/39) and *SLC6A1* 3.7% (6/160). Insufficient data is available to extrapolate the genetic contribution of *SYNGAP1, STX1B*, *TBC1D24* and *KIAA2022* in MAE, and I have estimated it at no more than 1 to 2% contribution each. The contribution of CNVs is not insignificant but appears most relevant when overlapping with epilepsy-associated genes.

This genetic heterogeneity may in part be accounted for by the phenotypic heterogeneity in MAE, as demonstrated by phenotypic clues in MAE cases for specific genes, e.g. movement disorders in *SLC2A1* and febrile seizures in *STX1B.* Phenotypic heterogeneity is compounded by the difficulties in a precise epilepsy syndrome diagnosis due to the clinical continuum of MAE and the intricacies in obtaining accurate seizure types. Hence a project incorporating deep phenotyping using well-validated instruments along with an NGS platform to generate sequence data was conceived. Exome sequencing was considered in preference to whole genome sequencing due to a more favourable cost versus output balance.

Exome sequencing is a particularly practical strategy to interrogate annotated protein coding genomic regions at both a reduced cost and arguably a more tailored dataset encompassing only 1.5% of the human genome where genetic mutations responsible for mendelian disorders are thought to lie. This has been particularly successful in rare mendelian disorders but also in epilepsy[20,77]. A good early example of the use of exome sequencing was in Kabuki syndrome[97]. The authors sequenced 10 phenotypically ranked cases based on the subjective assessment of the presence of canonical facial characteristics of Kabuki syndrome. *MLL2* mutations were identified in seven cases. They identified two further mutation positive cases by Sanger sequencing, and a less carefully selected replication cohort of 43 identified 26 mutation positive cases. This publication highlights that (1) Sanger sequencing is the gold standard for validating mutation identification following exome sequencing, (2) deep phenotyping can be useful to isolate a homogenous cohort and (3) phenotypic heterogeneity can reduce the

likelihood of gene identification as demonstrated by the reduced identification in the less carefully selected replicated cohort.

Disease identification strategies using exome sequencing can be divided into four main groups[98], all of which have been employed successfully in epilepsy. (1) Studies that use genetic linkage[99] (2) studies that employ homozygosity or a recessive model of inheritance in analysis[100] (3) studies that seek to identify overlapping shared rare variants or *de novo* variants[101] and lastly (4) studies that use a candidate gene based strategy[12]. These exome sequencing strategies are particularly effective in dealing with rare monogenic disorders but complex genetically heterogenous disorders often require more deciphering.

The high prevalence of diverse seizure and EEG traits within families of MAE probands suggest a more complex mode of inheritance where a combination of independent genetic factors give rise to different permutations of seizure and EEG features within individual family members. This may be analogous to the proposed model of oligogenic inheritance in GGE sub syndromes in which there are clear shared and distinct genetic influences on seizures and EEG[62,63,67,102,103]. To help unravel this, replicating family EEG studies to identify familial EEG traits may help in informing inheritance models. In addition specific EEG features such as GSW and PPR can be employed as endophenotypes for genetic analysis. Furthermore, as the spectrum of MAE extends to EE, this also favours a collateral approach using NGS to find monogenic causes.

My project seeks to close three major gaps in our knowledge of MAE. First, I shall investigate the neurodevelopmental phenotype, specifically seeking to identify deficits in cognition, socio-communication, attention deficit hyperactivity traits and adaptive functioning. Second, I shall perform family EEG studies to identify heritable EEG traits and potentially gain insight into patterns of inheritance. Third, I shall use exome sequencing to identify causative variants associated with MAE.

1.9 Project Aims and Objectives

This project's aims and objectives can be divided into three main arms. I have detailed which are the relevant chapters for each arm.

Chapters 2 and 3

1. I shall investigate the neurodevelopmental phenotype of 100 MAE cases.
    a. Behavioural comorbidity (autism spectrum, attention deficit hyperactive disorder) and adaptive behaviour will be investigated with the Social communication questionnaire (SCQ), Strength and difficulties questionnaire (SDQ), developmental, dimensional and diagnostic interview (3di) and the Adaptive behaviour assessment system (ABAS).
    b. Subjects aged between 3 and 7 years will be directly phenotyped for cognition using the Bayley scales of infant and toddler development (Bayleys III) or Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III); and language using the Clinical Evaluation of Language Fundamentals – Preschool (Pre School CELF) or Bayleys III.
    c. Detailed electroclinical features will be obtained. Participants will also complete a clinical questionnaire, ethnicity form and three-generation pedigree.

Chapters 4 and 5

2. I shall perform resting and sleep EEG studies in first degree relatives of 20 MAE families to identify generalised spike wave, photosensitivity and background rhythms and score for seizure type using ILAE classification.
    a. EEG trait data will be compared with population prevalence.
    b. I shall analyse concordance of EEG patterns amongst parents and siblings using permutation testing to assess significance.

Chapters 6 and 7

3. I shall isolate a causative genetic variant by collecting DNA from 100 MAE cases for exome sequencing.
    a. Different gene filtering methods are applied to navigate the remaining exome space in light of the rich genetic heterogeneity of MAE.

Chapter 2 Methods: Recruitment and Phenotyping

2.1 Study setup

This study is under the umbrella study of Professor Deb Pal's study in Genetics of Human Epilepsy, NIHR portfolio study 8858, CSP 34723. The protocol, participant information sheet and consent form was revised to include the relevant arms of this study. This was submitted and approved in stages to the Bloomsbury research ethics committee (ref:09/H0713/76). The versions and date-controlled documents used for the study were as follows: Genetics in Human epilepsy protocol, version 6.3, 25th November 2013, Adult and Parent of Minor participant information sheet and consent form, version 4, 5th December 2013 (see appendix A).

2.2 Participant recruitment

The case definition of MAE used was based upon the ILAE classification[29]. Cases were recruited if they met the following criteria, (1) usually normal development before onset of epilepsy; (2) onset of myoclonic, myoclonic-astatic or astatic seizures between 7 months and 6 years of age; (3) presence of generalise spike wave and/or polyspike wave discharges on EEG and/or absence of EEG indicating other epilepsy syndrome and (4) absence of related structural cerebral abnormalities on MRI.

MAE is a rare childhood epilepsy. The incidence of MAE is generally low and accounts for 1 – 5% of childhood epilepsies in different centres[30,33,37]. Due to the rarity of the disease, the study was set up in multiple sites in the UK and Europe in order to achieve recruitment targets. UK and International collaborators were assembled resulting in the collection of MAE cases, which were divided into three main cohorts, the UK, Euroepinomics and Italian cohorts. Research and Development approval was obtained for each site.

Additionally an internet based presence for the study was set up to raise awareness for the study and prompt referrals. The study was publicised in the laboratory group website www.childhood-epilepsy.org; in the UK epilepsy action website www.epilepsy.org.uk/info/syndromes/myoclonic-astatic-epilepsy-doose-syndrome, circulated amongst members of the patient groups Doose Syndrome Epilepsy Alliance www.doosesyndrome.org, and the facebook group Doose Syndrome U.K..

2.3 Collaborators

2.3.1 UK Clinical Collaborators

- Dr Elaine Hughes, Dr Karine Lascelles, Dr Ruth Williams, Dr Sushma Goyal, Evelina Children's Hospital, London.
- Professor Helen Cross, Dr Christin Eltze, Dr Robert Robinson, Institute of Child's Health, Great Ormond Street Hospital, London.
- Dr Penny Fallon, Dr Tim Kerr, St George's Hospital, London.
- Dr Maria Kinali, Chelsea and Westminister Hospital, London.
- Dr Alasdair Parker, Addenbrookes Hospital, Cambridge.
- Dr Shakti Agrawal, Birmingham Children's Hospital, Birmingham.
- Dr Nahin Hussain, University Hospital of Leicester, Leicestershire.
- Dr William Whitehouse, Nottingham University Hospital, Nottingham.
- Dr Colin Ferrie, Leeds Teaching Hospital NHS Trust, Leeds.
- Dr Archana Desurka, Sheffield Universtiy Hospital, Sheffield.
- Dr Siobhan West, Dr Grace Vassallo, Manchester Children's Hospital, Manchester.

2.3.2 International Clinical Collaborators

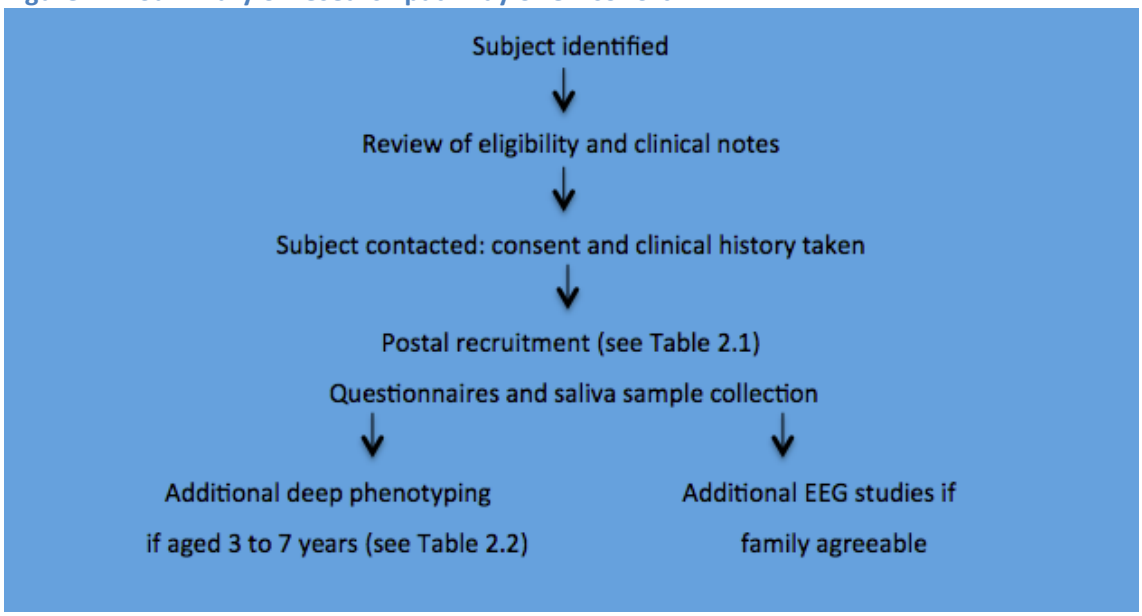
- Euroepinomics Rare Epilepsy Syndrome MAE working group. The members are Carla Marini, Renzo Guerrini, Bernd Neubauer, Christian M. Korff, Dana Craiu, Deb Pal, Hande Caglayan, Ingo Helbig, Peter De Jonghe, Rhys Thomas, Rikke S. Møller, Sanjay Sisodiya, Sarah von Spiczak, Sarah Weckhuysen, Tiina Talvik.
- Professor Renzo Guerrini, Professor Carla Marini, Anna Meyer Children's Hospital, Florence, Italy
- Dr Roberto Caraballo, Paediatric Hospital Dr Juan P Garrahan, Bueno Aires, Argentina.

2.4 UK cohort: Recruitment and phenotyping pathway

The UK cohort was made up of cases that were either referred from UK clinical collaborators or families self referred primarily through website enquiry. Medical and EEG reports were obtained from the referring physician and/or family. The clinical history for all referred cases was reviewed to determine eligibility. Eligible cases were then contacted either in person or over the phone where a clinical history and family history was taken. The study was discussed and consent was taken from the parent/caregiver. Cases were then stratified into either remote phenotyping through postal questionnaires only or with additional deep phenotyping

in research clinics. Families were also invited to attend King's Health Partners for EEG studies. See Figure 2.1 for summary of the research pathway for the UK cohort.

Figure 2.1. Summary of research pathway of UK cohort.



A research pack containing participant information sheets, consent forms, questionnaires for the parent/caregiver to complete and clearly labelled DNA and RNA saliva kits was sent to the families along with a prepaid addressed envelope to return the research materials. Participants were reminded at two to four weekly intervals to return the research packs. Table 2.1 summarises the phenotyping instruments used for all participants.

Table 2.1 Phenotyping instruments for all participants.

Instrument	**Age**	**Mode of administration**
Medical History	all	Over the telephone
Family History and Pedigree	all	Over the telephone
Child Form C	all	Postal questionnaire
Ethnicity Form	all	Postal questionnaire
Adaptive Behaviour Assessment System	all	Postal questionnaire
Strength and Difficulties Questionnaire	3 – 16 years	Postal questionnaire
Social Communication Questionnaire	> 4 years	Postal questionnaire
Parents and Teachers Conner's Rating Scale	> 4 years	Postal questionnaire

2.5 Deep Phenotyping

All subjects were invited for deep phenotyping with family EEG studies (see Chapter 4) and interview with the 3di (see section 2.6.4) and in subjects aged 3 to 7 years for additional deep phenotyping of cognitive and language abilities.

The age range for assessment of cognitive and language abilities was selected to capture most of the children with active epilepsy and under treatment as the mean age of seizure onset is between 32 to 43 months and approximately two thirds enter seizure cessation by 42 months [33,34,57]. Subjects aged between 3 and 7 years were invited for cognitive testing with the WPSSI III or Bayleys III and language testing with the Preschool CELF directly at sites in King's Health Partners in London or at their home if they lived close to London.

Research clinics were set up at three sites: (1) The Newcomen centre at St Thomas' Hospital, (2) The MRC Social Genetic Developmental Psychiatry (SGDP) building at the Institute of Psychiatry, Psychology and Neuroscience and (3) The Clinical Research Facility at King's College Hospital. Table 2.2 summarises the protocol for additional deep phenotyping at research clinics or subject's home.

Table 2.2 Protocol for deep phenotyping.

Instrument	Typical Age	Duration	Person
Preschool CELF	3 – 7 years	30 minutes	Investigator + case
Baileys III	6 – 42 months	30 – 90 minutes	Investigator + case
WPSSI III	3 – 7 years	60 minutes	Investigator + case
3di	all	30 – 45 minutes	Investigator + parent
Total duration	**150 – 225 minutes**		

2.6 Phenotyping Instruments for the UK cohort

2.6.1 General medical information

Medical notes including details of seizure onset, seizure types and EEG reports were obtained from referring clinical collaborators. Clinical history was also taken over the telephone or directly from all families. In addition, Form C, a questionnaire relating to the child's medical history and an ethnicity form generated from the ethnicity groups in the National census 2011 were given to parents to complete (appendix B). An. Family history was obtained from referring clinicians and directly from participants.

2.6.2 Cognition

The Wechsler Preschool and Primary Scale of Intelligence III UK (WPPSI III) and Bayley Scale of Infant and Toddler Development 3rd edition (Bayleys III) were used for cognitive testing. Both instruments are well-validated cognitive tests and are used frequently in research and in clinical practice[104]. The instrument selected was based on the estimated cognitive ability of the subject. Normative data for both instruments are available from their respective technical manuals.

The Wechsler Preschool and Primary Scale of Intelligence III UK

The WPPSI III is the most commonly used measure for intelligence of children from 2 years 6 months to 7 years 3 months. Separate subtests are administered to subjects above and below 4 years. Four core subtests were administered to subjects below 4 years. They are Receptive vocabulary and Information subtest for verbal IQ, and the Block design and Object Assembly for performance IQ. Subjects above 4 years were administered seven core subtests. The core verbal subtests are Information, Vocabulary and Word reasoning and the core performance subtests are Block design, Matrix reasoning, Picture concepts and Coding. A full scale IQ with a verbal and performance cognitive domain and/or developmental age equivalent was obtained from testing.

Bayley Scale of Infant and Toddler Development 3rd edition (Bayleys III)

The Bayleys III is designed to assess the development of infants and toddlers across cognitive, motor and developmental stages[105]. It was used when subjects had a cognitive developmental age and/or language ability below the age of 3 years and hence were unable to be tested with the WPPSI III and CELF Preschool. The cognitive scale of the Bayleys III measure function similar to the WPPSI III and the receptive communication and expressive communication subtest were used to obtain a language score. An age equivalent and/or scaled score was obtained from testing.

2.6.3 Language

The Clinical Evaluation of Language Fundamentals Preschool 2nd UK edition

The Clinical Evaluation of Language Fundamentals Preschool 2nd UK edition (CELF Preschool 2) measures expressive and receptive language and derives a core language index and percentile score in children aged 3 to 7 years. Subtest in sentence structure, word structure and expressive vocabulary were administered to participants. The core language score is obtained from the results of these three subtests.

2.6.4 Autism spectrum disorder

Autism spectrum disorder (ASD) refers to a group of developmental disorders characterised by a wide range of impairments in social and communicative abilities, stereotyped behaviours, and restricted range of interest with onset of initial symptoms before 3 years[106]. ASD was screened using the Social Communication Questionnaire (SCQ) and where possible symptoms were also elicited directly with a parent/care giver with the 3di.

The Social Communication questionnaire

The Social Communication questionnaire (SCQ) is a brief parent-report diagnostic screening measure that helps to evaluate communications skills and social functioning in order to screen for Autism[107]. It is suitable for children aged above 4 years with a mental age above 2 years. The SCQ was designed based on a selection of questions from the Autism Diagnostic Interview – Revised, an algorithm used for Diagnostic and Statistical Manual of Mental Disorders (DSM) IV diagnosis of autism. The DSM serves as an authority for classification and diagnosis for psychiatric diagnosis[108].

The SCQ Autoscore Form: Lifetime was used. The form references the individual's behaviour during the individual's entire developmental history. There are 12 questions that cover three areas of functioning: reciprocal social interaction, communication and restricted repetitive and stereotyped behaviour. Particular behaviour with associated statements was rated as 'not true', 'quite or sometimes true' or 'very or often true' with scores of 0-1-2. The threshold for suspecting a social communication disorder was a score of 15 and over.

The Developmental, Dimensional and Diagnostic Interview

The Developmental, Dimensional and Diagnostic Interview (3di) is a computerized interview assessment that identifies clinical and subclinical ASD[109]. The 3di scores for reciprocal social interaction skills, use of language and other social communication skills and repetitive/stereotyped behaviours and routines. The full 3di is 117 questions and requires 2 hours so the ASD rapid assessment comprising 53 questions was used. The 53 questions in the ASD rapid assessment was found to be most discriminating and can be scaled up to the same diagnostic thresholds as the DSM-5 for ASD[110]. The reduced testing burden makes it more feasible to be administered to participants.

3di diagnosis are made in general on the following criteria where A is social reciprocity, B communication, Bnv non verbal communication and C restricted/repetitive behaviours and interest:

Child 24 - 35 months or > 36 months with no speech, single words or phrase speech

Autism – meets criteria for A and Bnv and C

Atypical autism – meets criteria for A and either Bnv or C

Child with fluent speech or Child > 36 months but < 48 months with phrase speech

Autism – meets criteria for A and B and C

Asperger – meets criteria for A and B and C with no developmental problems in the onset of language

Atypical autism – meets criteria for A and either B or C

2.6.5 The Strength and Difficulties questionnaire

The Strength and Difficulties questionnaire (SDQ) is a brief behavioural psychosocial screening questionnaire for 3-16 year olds[111]. It is well validated as a first pass screening psychiatric questionnaire and developed from the Rutter's questionnaire, but designed to be more brief than the Rutter's questionnaire and Child Behaviour Checklist[112].

Two versions of the SDQ were used. The parent 4 to 16 year old version and a slightly modified informant rated version for 3 to 4 year olds. There are 25 items to ascertain positive or negative attribute in five scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. The items can be rated "not true", "somewhat true" or "certainly true" with scores 0-1-2. An impact supplement is also included which enquires further about chronicity, distress, social impairment and burden to others. The scores can then be classified into close to average, slightly raised, high or very high.

2.6.6 The Conner's Comprehensive Behaviour Rating Scale

Attention Deficit Hyperactivity Disorder (ADHD) is a childhood-onset neurodevelopmental disorder characterised by deficits in attention and/or hyperactivity and impulsivity. The Conner's Comprehensive Behaviour Rating Scales (CBRS) was used as the assessment tool to identify relevant symptoms.

The CBRS is a comprehensive assessment tool, which assesses a wide range of behavioural, emotional and social disorders[113]. Parents and the child's teacher are asked to complete separate questionnaires. The subscales are based on the symptom subscales of inattentive,

hyperactive-impulsive as well as other symptoms of the DSM IV[108]. A summary of the different subscales is shown in appendix C.

The raw scores are converted to a gender and age specific T score across different subscales. A T score is a standardised score with a mean of 50 and a standard deviation of 10. A T score of 70 and above corresponds to 98th percentile and above and was used as a cut-off for significant difficulties in the individual subscales and for meeting DSM criteria of ADHD.

2.6.7 The Adaptive Behaviour Assessment System

The Adaptive Behaviour Assessment System (ABAS II) provides a questionnaire assessment of daily adaptive skills for functioning effectively given the typical demands placed on a child[114]. The importance of adaptive behaviour in daily activities in the context of intellectual disability is increasingly recognised.

In 2002, the American Association on Intellectual and Developmental Disabilities defined disability as characterised by significant limitation in both intellectual functioning and adaptive behaviour including conceptual, social and practical skills. In the UK, the Equality Act 2010 defined disability as a physical or mental impairment that has a substantial and long-term negative effect on the ability to do normal daily activities. Moreover, the recent DSM V diagnostic criteria for intellectual disability now includes criterion for deficits in adaptive functioning.

The ABAS II provides overall adaptive functioning (General Adaptive Composite) as well as assessment of the ten adaptive skill areas specified in the DSM IV. Additionally it groups skill areas into three adaptive domains: conceptual, social and practical (See Table 2.3). The two forms that were used were the Parent Form for birth to 5 years that includes 241 items and the Parent Form for 5 to 21 years, which includes 232 items.

Table 2.3 Adaptive domains explored in the ABAS.

Adaptive Domain	Skill area
General Adaptive Composite	Communication Community Use Functional Pre-Academics Home Living Health and Safety Leisure Self-Care Self-Direction Social Motor/Work*
Conceptual	Communication Functional Pre-Academics Self-Direction
Social	Leisure Social
Practical	Self-Care Home Living Community Use Health and Safety

* Motor for birth – 5 years and Work for 5 – 21 years

The results are represented as standardised data and allow a comparison between an individual's adaptive skills and typically developing individuals of the same age. A General Adaptive Composite (GAC) score was derived from the sum of the conceptual, social and practical domains. These scores were then classified into very superior (≥130 score, ≥98th percentile), superior (120-129 score, 91st-97th percentile), above average (110-119 score, 75th-90th percentile), average (90-109 score, 25th-74th percentile), below average (80-89 score, 9-24th percentile), borderline (71-79 score, 3rd-8th percentile) and extremely low (≤70 score, ≤2nd percentile) as had been practically done by other studies[115].

2.7 Data collection for the Euroepinomics and Italian cohort

Phenotypic data of the cases from the Euroepinomics cohort were available through an on-line password protected platform http://redcap.uni.lu. Where data were missing, referring clinicians were contacted directly for further clarification. Phenotypic data of the Italian cohort were relayed directly from the Italian collaborators. Specific details regarding seizure onset, seizure types, EEG features and comorbidity such as ID, ASD and ADHD in cases were collected.

Chapter 3 Results: The epilepsy and neurodevelopmental phenotype of MAE

3.1 MAE cohorts

123 MAE cases were collected from three main cohorts. These were (1) 67 cases from the UK cohort, (2) 38 cases from the Euroepinomics consortium cohort and (3) 18 cases from the Italian cohort. The ethnicity make up of the different cohorts is shown in Table 3.1.

Table 3.1: Ethnic groups of the three MAE cohorts

	UK cohort	Euroepinomics cohort	Italian cohort	Total
Ethnic groups				
White				
White British	40	0	0	40
Other White	14	35	16	65
Black African or Caribbean	4	0	0	4
Mixed ethnic group	5	1	2	8
Asian	4	1	0	5
Unknown	0	1	0	1
Total	**67**	**38**	**18**	**123**

White British consisted of White English, Welsh, Scottish or Northern Irish.
Other White consisted of other White European or White American.
Mixed ethnic group consisted of White and Black African/Caribbean or White and Asian groups.
Asian consisted of South Asian, Algerian.

3.2 Epilepsy phenotype

There were 83 males (67.4%) and 40 females (32.5%) in the entire cohort. A family history of epilepsy was reported in 44 (37.6%) out of 117 cases. This was a first-degree family member in 16 (13.6%) cases and second or more degree family member in 28 (23.9%) cases. A family history of febrile seizures was reported in 10 (8.5%) cases. There were missing data on family history in five Italian cases (009Z, 565D, 561D, 224D, 027A) and one Euroepinomics case (W_M_1144326).

20 (17.5%) out of 114 cases reported developmental delay prior to epilepsy onset. Of these cases with developmental delay, nine cases reported isolated speech delay prior to seizure onset. Seven Italian cases (291J, 731J, 565D, 561D, 224D, 768H, 027A) and two cases (W_M_1158897, W_M_1144306) in the Euroepinomics cohort had missing information about

prior development. Thirty-four (27.6%) cases had a personal history of febrile convulsions. Table 3.2 summarises the characteristics of the three MAE cohorts.

Table 3.2. Summary of the main characteristics of the three MAE cohorts.

Cohort	**Boys (%)** (n=123)	**Febrile seizures (%)** (n=123)	**Family history Epilepsy, FS (%)** (n=117)	**Median age of onset in months (range)** (n=122)
UK (n=67)	49 (73.1%)	23 (34.3%)	32 (47.7%), 6 (8.9%)	35 (6-65)
Euroepinomics (n=38)	26 (68.4%)	3 (7.8%)	9 (24.3%), 4 (10.8%)	33 (3-60)
Italian (n=18)	8 (44.4%)	8 (44.4%)	3 (23.0%), 0	31.5 (6-72)
Total cohort (n=123)	83 (67.4%)	34 (27.6%)	44 (37.6%), 10 (8.5%)	35 (3-72)

FS febrile seizures

The median age of seizure onset was 35 months (range 3 to 72 months) based on 122 cases. Age of seizure onset data were missing for one Euroepinomics case (W_M_1158865). Seizure types at onset were GTCS in 63 (51.6%) cases, myoclonic atonic or atonic in 33 (27.0%) cases, myoclonic in 16 (13.1%) cases, absence in four (3.2%) cases and focal or tonic in three (2.4%) cases. Seizure type at onset was missing for one Euroepinomics case (EG0679). During the course of the epilepsy, mainly generalised seizure types were reported. 115 (93.4%) cases reported a history of myoclonic-atonic or atonic seizures, 85 (69.1%) cases reported GTCS, 84 (68.2%) cases reported myoclonic seizures, 67 (54.4%) cases reported absence seizures, 26 (21.1%) cases reported tonic seizures, eight (6.5%) cases reported focal seizures with or without secondary generalisation, four (3.2%) cases reported epileptic spasms and two (1.6%) cases had gelastic seizures. Table 3.3 summarises the main seizures types in the three MAE cohorts.

Table 3.3 Main seizure types in the three MAE cohorts.

Cohort	Myo-atonic or atonic	GTCS	Myoclonic	Absence	Tonic	Focal
UK (n=67)	67 (100%)	55 (82.0%)	56 (83.5%)	43 (64.1%)	15 (22.3%)	6 (8.9%)
Euroepinomics (n=38)	35 (92.1%)	21 (55.2%)	20 (52.6%)	16 (42.1%)	3 (7.8%)	2 (5.2%)
Italian (n=18)	13 (72.2%)	9 (50%)	8 (44.4%)	8 (44.4%)	8 (44.4%)	0
Total (n=123)	115 (93.4%)	85 (69.1%)	84 (68.2%)	67 (54.4%)	26 (21.1%)	8 (6.5%)

Clinical examination findings were available in 109 cases, and were reported abnormal in 23 (21.1%) cases. These were cerebellar signs mainly involving tremor and ataxia in 13 cases, six cases with pyramidal or motor signs, three cases with dysmorphism and one case with microcephaly.

EEG reports were available in 112 cases (61 UK cohort, 37 Euroepinomics cohort, 14 Italian cohort). The majority, 104 (92.8%) cases demonstrated generalised epileptiform activity of spike wave discharges. Additional EEG features in these cases are summarised in Table 3.4. In the remaining eight cases without generalised epileptiform activity, four had prominent theta rhythms on their interictal EEG, two had focal changes and two had a normal EEG.

Table 3.4 Additional reported EEG features in the 100 MAE cases with generalised epileptiform activity.

Additional EEG features	Number of cases
Polyspike and wave	27
Slow background	15
Focal epileptiform activity	11
Photoparyoxysmal responses	3
Hyperventilation activation	2

Seizure Remission

Data regarding seizure remission where seizure freedom had been achieved for more than two years with or without the presence of anti-epileptic drugs was obtained for the UK and Italian cohort. In the UK cohort, 20 (29.8%) cases were in seizure remission and in the Italian cohort seven (38.8%) cases were reported to be in seizure remission.

3.3 Intellectual disability and Language

Reported

Intellectual disability (ID) was a well-recognised co-morbidity. As a general measure, the presence of ID was elicited through data provided by the referring clinician. ID was reported in 43 (64.1%) cases in the UK cohort, 23 (60.5%) cases in the Euroepinomics cohort and 10 (55.5%) of the Italian cohort. Data from six cases (W_M_1147816/1151981/1151131/1158897/1199453/1202321) of the Euroepinomics cohort were missing. Taking into account missing data, ID was reported in 76 (64.9%) out of 117 cases. It is difficult to determine the exact extent of cognitive deficit without formal cognitive testing.

Formal testing - cognition

Formal cognitive testing results in 25 cases showed a breakdown of 56% with moderate to severe ID, 16% with mild ID and 28% as normal.

Formal cognitive testing in the UK cohort was carried out in 15 cases with the WPPSI III (n=8) and with the Bayleys III (n=7). In two cases, subject 00568 and subject 00596, testing was discontinued due to behavioural difficulties, as they were not able to comply with the assessments required. In addition, results from clinical cognitive testing were available from a further 12 cases. Table 3.5 and Table 3.6 detail these results.

ID was classified as moderate to severe where IQ <70, cognitive ability >2 SDs below the mean; mild where IQ 71 - 85, cognitive ability between 1 and 2 SDs below the mean, and normal, where IQ >85, within 1 SD of the mean. In the five cases that were tested with the Bayleys III, it was possible only to obtain cognitive scaled scores for two cases. A scaled score represents a child's performance relative to a same age peer. They are scaled to a metric with a range of 1 to 19, a mean of 10 and a SD of 3. Subject 00572 achieved a scaled score of one (~IQ 55) and subject 00574 a scaled score of six (~IQ 75). The other three cases had chronological age above

the testing age of the Bayleys III so it was not possible to attain a scaled score/percentile rank and only age equivalents were possible.

Formal testing - language

Language testing results were available in 14 cases, with the CELF-P or CELF (n=9) and Bayleys III (n=5). In total, there was evidence of delayed language in nine (64.2%) cases. Percentile ranks for the core language scores from the CELF is shown in Table 3.5. In four cases, this percentile rank was less than the 10th centile. Receptive and expressive language testing was also performed in five cases with the Bayleys III. In all five cases age equivalents in cognition, receptive language and expressive language were below their chronological age with language more severely affected than cognition (see Table 3.6). In the two cases where a scaled score was possible, subject 00572 attained a scaled score of two for receptive and three for expressive communication and subject 00574 attained a scaled score of two for both receptive and expressive communication.

Table 3.5: Results of IQ and Language tests.

Subject ID	Age of seizure onset	Age at testing	IQ assessment tool	VIQ	PIQ	FSIQ (95% C.I.)	CELF P/ CEF
3002 301*	4y 7m	7y 2m	WISC IV UK	96	98	96 (91-101)	-
3004 301	3y 7m	6y 2m	WPPSI III	101	73	80 (76-96)	8th
00505	2y	6y 5m	WPPSI III	102	77	90 (85-95)	45th
00506*	3y	6y 5m	WPPSI III	72	65	60 (56-66)	-
00524	4y	5y 8m	WPPSI III	104	112	110 (105-115)	25th
00526	2y 11m	5y 8m	WPPSI III	61	59	48 (45-54)	0.1th
00539	2y 6m	5y 10m	WPPSI III	95	90	93 (88-98)	66th
00566*	4y 3m	12y	WISC IV	69	77	70 (65-75)	-
00600*	2y 11m	4y 2m	WPPSI IV	75	75	67 (63-73)	<0.1th
00602	2y 11m	7y 8m	WPPSI III UK	101	119	99 (94-105)	6th
00608	3y 6m	8y 4m	WPPSI III UK	73	77	72 (68-78)	21st
00615	4y 8m	8y 0m	WPPSI III UK	100	100	97 (92-102)	50th
W_M_1147805*	3y 7m	9y 7m	WISC III	-	-	35- 49	-
W_M_1158865*	-	12y	WISC III	-	-	49 (46-55)	-
W_M_1152720*	3y	-	WISC IV	-	-	50 - 69	-
W_M_1145972*	1y 4m	4y 9m	WPPSI III	-	-	70 - 89	-
W_M_1195591*	4y	13y	-	-	-	60	-
W_M_1199464*	3y	17y	-	-	-	48	-
072G*	1y 9m	4y 10m	-	87	89	87	-

VIQ verbal IQ, PIQ performance IQ, FSIQ full scale IQ, *results obtained from clinical testing reports, - data not available, y years, m months

Table 3.6: Results of Bayleys III testing.

Subject ID	Age of seizure onset	Age at testing	Cognitive Age Equivalent	Receptive language age equivalent	Expressive language age equivalent
00518	2y 11m	6y 0m	27m	22m	17m
00560	2y 6m	5y 10m	7m	<1m	7m
00572	9m	2y 4m	10m	6m	12m
00574	2y 4m	2y 9m	22m	9m	10m
00590*	3y	6y 6m	23m	-	-
00599	9m	7y 3m	28m	19m	9m

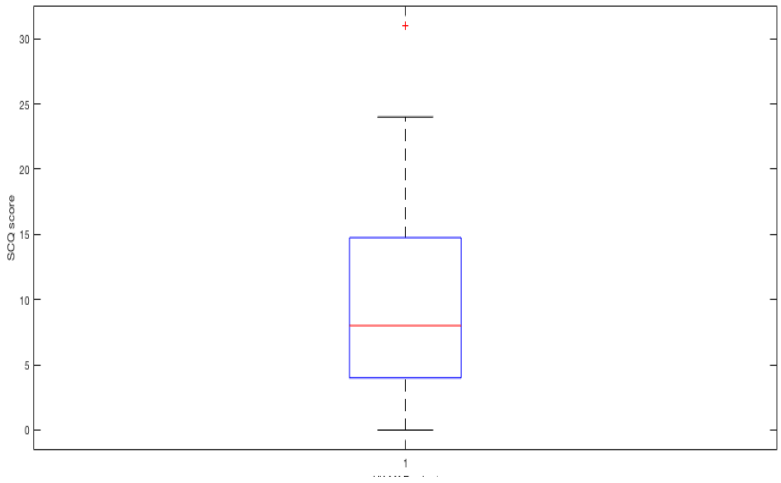
* Results obtained from clinical testing reports, y years, m months

3.4 Autism

Autistic spectrum disorder (ASD) features were elicited from clinician report in the Italian and Euroepinomics cohort and through the Social Communication Questionnaire (SCQ) and the dimensional and diagnostic interview (3di) in the UK MAE cohort.

Sixty-one (43 males, 18 females) of 67 SCQ questionnaires were returned in the UK cohort, achieving a response rate of 91.0%. Sixteen (26.2%) cases reached the threshold for suspecting a social communication disorder with a score of 15 and over. The mean score was 8.65, SD 7.32, median 8 (range 4 - 31). Figure 3.1 shows the box plot distribution of SCQ scores. Based on clinician report, two (20%) (660M, 768H) out of ten subjects in the Italian cohort and four (12.5%) (W_M_1145223,1158858,1151131,1145970) out of 32 subjects in the Euroepinomics cohort reported a diagnosis of Autistic spectrum disorder. Eight subjects in the Italian cohort and six subjects in the Euroepinomics cohort had missing data. Taking into account all these measures, 22 (21.3%) out of 103 subjects reported a diagnosis or symptoms of ASD

Figure 3.1 Boxplot of SCQ scores of the UK MAE cohort.



The SCQ scores were also distributed by gender and compared with the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort of 4167 boys and 3927 girls[116]. The ALSPAC cohort is a prospective longitudinal intensively studied population cohort of children. Initial recruitment targeted all pregnant women living in the geographical region of Avon, England who were expected to deliver their baby between April 1, 1991 and December 31, 1992. Contact has been maintained with the families and surviving children (http://www.bris.ac.uk/alspac/sci-com). In the ALSPAC cohort, males had significantly higher scores than females but this was not seen in the UK MAE cohort. However, the UK MAE cohort had significantly higher SCQ scores than the ALSPAC cohort in both males and females (*P*<0.0001) (see Table 3.7).

Table 3.7: Comparison of SCQ scores between males and females and with the ALSPAC cohort.

	Males mean (SD)	**Females mean (SD)**	***P* value for gender comparison**
UK MAE cohort	8.73 (7.37)	12.11 (7.13)	0.1045
ALSPAC	3.25 (4.15)	2.39 (3.14)	< 0.0001
P value for cohort comparison	< 0.0001	< 0.0001	

Two-tailed *P* value calculated with student t test.

Detailed ASD symptoms were investigated further with the 3di interview in the UK MAE cohort. Twenty-two 3di interviews with one or both parents were conducted and the results are shown in Table 3.8. Based on the results of the 3di interview, nine cases had an ASD diagnosis. The ASD diagnosis was Autism in three cases, Atypical autism in four cases and Asperger in two cases. Cases reported most difficulties in the communication domain with a mean score of 9.3 (SD 7.1) within the abnormal range for that subscale.

Comparison of the scores in individual domains were made with typically developing and ASD Thai chilren[117]. At the time of writing (February 2017), I am awaiting data for the 3di scores from British children and in the interim have used published data from Thai children. Scores in social reciprocity and restricted/repetitive behaviours and interest were significantly higher than typically developing (*P*=0.03) and yet significant less than ASD children (*P*=0.002). This suggests that the MAE children had specific subclinical ASD symptoms with significant levels of difficulties. As for scores in communication, the MAE cohort had significantly higher scores than typically developing children (*P*=0.003) but not as significantly different with the ASD children (*P*=0.05); suggesting a similar level of difficulty in communication in the sample studied as ASD children.

Table 3.8 Results of individual 3di scores.

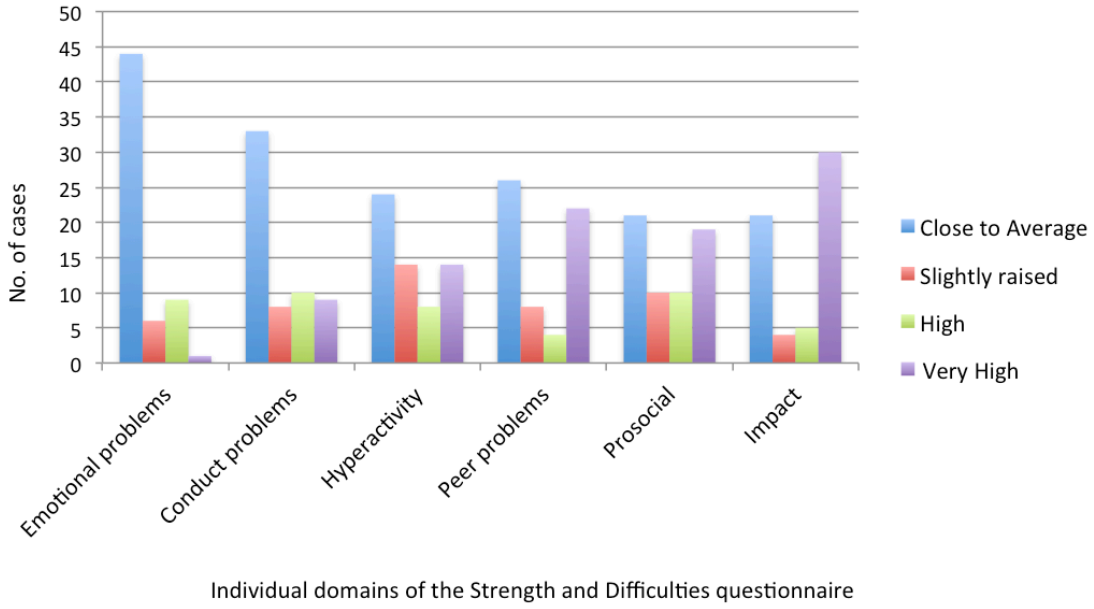
Subject ID	Age at interview	A: Social reciprocity [10-30]	B: Communication [8-26]	Bnv: Non-verbal [7-14]	C: RRBI [3-12]	3di diagnosis	SCQ score
3002 301	12y 5m	1	4.5	3	0	No ASD	0
3003 301	5y 5m	**11**	6	2	**4**	Atypical autism	-
3004 301	6y 2m	3.5	**9**	5	**8**	No ASD	8
00505	6y 5m	3.3	0.5	0	1	No ASD	1
00506	14y 6m	2	1.5	0	0	No ASD	7
00518	5y 11m	**20.3**	**12.5**	6	0	Atypical autism	14
00524	5y 7m	3.3	2.5	0.5	0	No ASD	4
00526	5y 8m	1.5	4.5	0.5	1	No ASD	2
00528	12y 3m	**16.8**	**14.2**	**7.7**	**7**	Asperger	**16**
00539	5y 10m	1	3.5	3.5	0	No ASD	0
00554	17y 5m	0	2	2	0	No ASD	0
00560	5y 10m	**18.7**	**15.3**	**11.3**	**5**	Autism	**18**
00568	6y 10m	**15.1**	**20.4**	**10.4**	**6**	Autism	-
00571	8y 8m	9.7	**10.8**	**8.8**	1	No ASD	9
00572	2y 4m	**13.8**	**17.8**	**9.8**	2	Atypical autism	-
00574	2y 10m	**14.3**	**9.6**	**7.6**	0	Atypical autism	10
00586	8y 9m	**19.8**	**15.7**	**8.2**	**4**	Autism	**16**
00591	6y 0m	5.7	3	2	0	No ASD	4
00599	7y 3m	**15.8**	**19**	**13**	**3**	Asperger	**24**
00602	7y 8m	3	3	2	1	No ASD	7
00611	9y 5m	**20.6**	**24.6**	**14**	2	No ASD	-
00615	8y 0m	6.7	5.5	3	2	No ASD	13
Median (range) MAE cohort		8.2 (0-20.6)	7.5 (0.5-24.6)	4.2 (0-14)	1 (0-8)		
Mean (SD) MAE cohort		9.4 (7.3)	**9.3** (7.1)	5.4 (4.4)	2.1 (2.4)		
Typical children[117] (n-67)		7.0 (3.1)	6.0 (3.1)	-	1.7 (1.2)		
ASD children[117] (n=63)		13.4 (4.1)	11.6 (3.7)	-	3.5 (1.9)		
Comparison of MAE cohort and typical children		***P* = 0.03**	***P* = 0.003**	-	***P* = 0.30**		
Comparison of MAE cohort and ASD children		***P* = 0.002**	*P* = 0.05	-	***P* = 0.006**		

[] Abnormal range, RRBI restricted/repetitive behaviours and interest, - no data, Unpaired student t test used to calculate levels of significance.

3.5 Strength and Difficulties Questionnaire

Sixty of 67 Strength and Difficulties questionnaires were returned, achieving a response rate of 89.5%. The raw scores in individual subscales were divided into close to average, slightly raised, high and very high and are represented in Figure 3.2. Cases reported high and very high scores across all domains. This consisted of ten (16.6%) cases with emotional problems, 19 (31.6%) cases with conduct problems, 22 (36.6%) cases with hyperactivity problems, 26 (43.3%) cases with peer problems, 29 (48.3%) cases with prosocial problems and 35 (58.3%) cases reporting these problems to significantly impact the family and child.

Figure 3.2 Individual domain scores of the Strength and Difficulties questionnaire.



The summary of the proportion of cases with high and very high scores for the various domains are demonstrated in Table 3.9, along with a normative sample of 10298 5 to 15 year old British children[118]. Scores in conduct problems (P<0.0001), hyperactivity/inattention (P<0.0001), peer relationship problems (P<0.0001), prosocial behaviour (P<0.0001), total difficulties (P<0.0001) and impact scores (P<0.0001) were significantly higher in cases compared with this normative British sample.

Table 3.9 Comparison of Strength and difficulties questionnaire scores with British children.

	Emotional symptoms (%)	Conduct problems (%)	Hyperactivity / inattention (%)	Peer relationship problems (%)	Prosocial behaviour (%)	Total difficulties score (%)	Impact score (%)
Proportion of cases with high/very high scores (n=60)	16.6	31.6	36.6	43.3	48.3	38.3	58.3
Proportion of British children with high/very high scores[118] (n=10298)	11.4	12.7	9.6	11.7	10.6	9.8	8.8
Level of significance (*P*)	0.206	**<0.0001**	**<0.0001**	**< 0.0001**	**< 0.0001**	**<0.0001**	**< 0.0001**

P determined by chi squared test

3.6 Conner's Behavioural Rating Scale and attention deficit hyperactivity disorder

ADHD symptoms were ascertained through clinician reports in the Italian and Euroepinomics cohorts and through the Conner's Parent and Teacher Behavioural Rating Scale (CBRS) with the UK cohort. Based on clinician report, three (30%) out of ten subjects in the Italian cohort, 15 (45.4%) cases out of 33 subjects in the Euroepinomics cohort reported ADHD symptoms, and 21 (40.3%) of 52 UK parents met CBRS questionnaire criteria of ADHD (subscale N, T score >70). This puts the total rate of ADHD symptoms in the entire cohort to 39 (41.0%) out of 95 cases.

Fifty-two out of 67 Parent CBRS questionnaires were returned in the UK cohort. The reduced response rate of 77.6% compared to other questionnaires is possibly due to the increased number of questions in this questionnaire. Forty-three Teacher CBRS questionnaires were collected achieving a response rate of 64.1%. The Parent and Teacher CBRS questionnaire have similar subscales apart from subscale G; the psychosomatic subscale that is absent in the Teacher questionnaire. Raw scores for each subscale are converted into a T score. Cases were then grouped into T scores groups of 40-59, 60-64, 65-69 and 70+. Number and proportion of cases in each subscale that fall into these T score groups are shown in Table 3.10.

The two most common T score groups in every subgroup were either scores below 59 or above 70; hence both parents and teachers either reported a symptom as being of concern or not. The most frequently reported concern for both parents and teachers independently and jointly were cognitive problems/inattention (subscale B), with 50.0% of parents and 34.8% of teachers scoring a T score >70.

Joint scoring of Parent and Teacher CBRS questionnaires and statistical comparison with a normative sample was also performed. For each subscale, the number of cases were assembled where T scores was 70 and above for both Parent and Teacher CBRS for that individual. Comparison was performed with the normative sample (n=3400) used in the development of the CBRS. A T score of 70 has a theoretical percentile of 97.7^{th} i.e. approximately 78 individuals from the normative sample had a score of 70 and above. Fisher's exact test was performed for each subscale, a multiple testing corrected level of significance for the ten subscales was set at $P<0.005$. This robust analysis demonstrated significant levels of cognitive problems (subscale B, $P=0.0001$), inattentiveness (subscale L, $P=0.0001$), restless impulsivity (subscale I, $P=0.0034$), oppositional (subscale A, $P=0.0005$) and ADHD (subscale N, $P=0.0001$) in the UK MAE cohort.

Table 3.10 Distribution of T scores in parent and teachers CBRS questionnaires.

T scores	40-59	60-64	65-69	70+	Joint T>70	Sig level
A: Oppositional	28 (53.8%)	4 (7.6%)	5 (9.6%)	15 (28.8%)	6	**0.0005**
	28 (65.1%)	3 (6.9%)	1 (2.3%)	11 (25.5%)		
B: Cognitive problems/ inattention	19 (36.5%)	5 (9.6%)	2 (3.8%)	26 (50.0%)	12	**0.0001**
	15 (34.8%)	6 (13.9%)	7 (16.2%)	15 (34.8%)		
C: Hyperactivity	23 (44.2%)	10 (19.2%)	3 (5.7%)	16 (30.7%)	3	0.0791
	30 (69.7%)	3 (6.9%)	5 (11.6%)	5 (11.6%)		
D: Anxious-Shy	39 (75.0%)	4 (7.6%)	4 (7.6%)	5 (9.6%)	0	1.0000
	23 (53.4%)	7 (16.2%)	6 (13.9%)	7 (16.2%)		
E: Perfectionism	39 (75.0%)	5 (9.6%)	2 (3.8%)	6 (11.5%)	0	1.0000
	37 (86.0%)	1 (2.3%)	3 (6.9%)	2 (4.6%)		
F: Social Problems	28 (53.8%)	3 (5.7%)	1 (1.9%)	20 (38.4%)	3	0.0791
	30 (69.7%)	1 (2.3%)	7 (16.2%)	5 (11.6%)		
G: Psychosomatic	27 (51.9%)	5 (9.6%)	7 (13.4%)	14 (26.9%)	-	**0.0001**
H: Conner's ADHD Index	19 (36.5%)	5 (9.6%)	4 (7.6%)	24 (46.1%)	9	**0.0001**
	23 (53.4%)	7 (16.2%)	4 (9.3%)	9 (20.9%)		
I: CGI Restless-Impulsive	27 (51.9%)	5 (9.6%)	2 (3.8%)	18 (34.6%)	5	**0.0034**
	24 (55.8%)	7 (16.2%)	4 (9.3%)	8 (18.6%)		
J: CGI Emotional Liability	27 (51.9%)	6 (11.5%)	6 (11.5%)	13 (25.0%)	3	0.0791
	29 (67.4%)	2 (4.6%)	3 (6.9%)	9 (20.9%)		
K: CGI Total	21 (40.3%)	7 (13.4%)	5 (9.6%)	19 (36.5%)	6	**0.0005**
	25 (58.1%)	4 (9.3%)	2 (4.6%)	12 (27.9%)		
L: DSM IV Inattentive	22 (42.3%)	2 (3.8%)	6 (11.5%)	22 (42.3%)	9	**0.0001**
	18 (41.8%)	8 (18.6%)	4 (7.3%)	13 (30.2%)		
M: DSM IV Hyperactive-Impulsive	23 (44.2%)	5 (9.6%)	5 (9.6%)	19 (36.5%)	4	0.0182
	28 (65.1%)	7 (13.4%)	3 (5.7%)	5 (9.6%)		
N: DSM IV ADHD	19 (36.5%)	3 (5.7%)	9 (17.3%)	21 (40.3%)	7	**0.0001**
	23 (53.4%)	3 (6.9%)	8 (18.6%)	9 (20.9%)		

Shaded cells are parents' questionnaires, ADHD attention deficit hyperactivity disorder, CGI Conner's global index, DSM Diagnostic and statistical manual of mental disorders, Sig level significance level *P*<0.005. Descriptions of the different subscales are in appendix C.

3.7 Adaptive behaviour

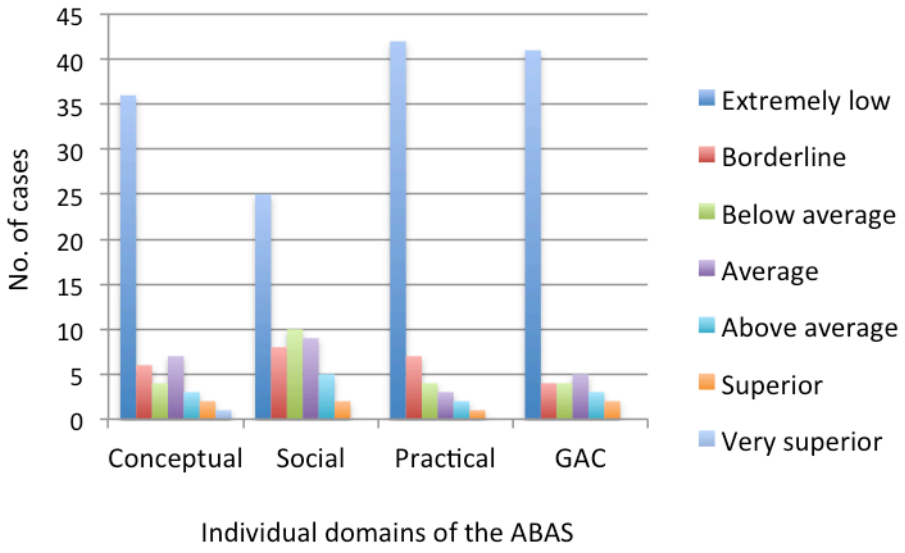
Adaptive behaviour skills were ascertained using the ABAS II questionnaire and were extremely low across all domains.

Fifty-nine of 67 ABAS II questionnaires in the UK cohort were returned, achieving a response rate of 88.0%. A composite score and percentile rank was obtained for each of the main ABAS domains of conceptual, social and practical. A general adaptive composite score was derived from the sum of these three domains.

Adaptive skills were low in all three domains of conceptual, social and practical and in the general adaptive composite score. Extremely low adaptive scores which indicated a percentile rank of ≤2 were the most commonly reported group and seen in 36 (61.0%) cases for conceptual, 25 (42.3%) cases for social and 42 (71.1%) cases for practical domains.

Correspondingly, 41 (69.4%) cases had extremely low scores for general adaptive composite score. Practical adaptive scores were the worst performing with less than average scores obtained in 53 (89.8%) compared with conceptual skills at 46 (77.9%) and social skills at 43 (72.8%) (see Figure 3.3)

Figure 3.3 Distribution of individual domain scores of the ABAS.



GAC general adaptive composite score

3.8 Other behavioural symptoms

During interview with the UK cohort and through clinical report, other behavioural concerns were identified. Ten families reported aggressive or violent behaviour in their children, out of which three children were on Risperidone. Risperidone is an antipsychotic medication that is used to treat schizophrenia and symptoms of bipolar disorder. Three children were reported to have sleep difficulties, one case had an obsessive-compulsive behaviour and one case was reported to have pain insensitivity.

Chapter 4 Methods: Electroencephalogram

4.1 Basic principles of the EEG

The electroencephlogram (EEG) endures as the first line investigative tool for supporting a diagnosis of epilepsy since Hans Berger's first publication of the alpha rhythm in 1929. The EEG has evolved from a clinical diagnostic tool in epilepsy into a computational analysis tool for many other neurological and psychiatric disorders. The principles of EEG recordings remain the same, independent of its final application.

The basis of the EEG is the recording of summated and synchronised postsynaptic field potentials of many cortical neurons recorded by multiple electrodes usually placed on the scalp surface. These electrical discharges are generated by excitatory postsynaptic potentials from apical dendrites of pyramidal cells on layer five of the cortex. Electrical activity is measured as the voltage difference between two electrodes. The electrodes are connected to a differential amplifier. The differential amplifier amplifies the difference in voltage between two electrode sites and also rejects the interference that simultaneously affects both electrodes. The combination of two electrodes in one amplifier is known as a derivation or channel and a montage consist of a combination of derivations. In a bipolar montage, each channel represents the potential difference between 2 adjacent electrodes in a chain. In a referential montage, each channel represents the potential difference between individual scalp electrodes and a common reference electrode. Figure 4.1 shows an example of a EEG system.

Figure 4.1 Nihon Koden EEG system

The waveforms recorded by the EEG scalp electrodes depend on the orientation and distance of the discharging neuron with respect to the recording electrode. Activity is maximal when discharges are orientated perpendicular to the scalp surface. The amplitude and morphology of waveforms can vary based on the proximity from the source to the recording electrode and follow the laws of volume conduction principles. Volume conduction is the process of current flow through the tissues (e.g. scalp, bone, cerebrospinal fluid) between the electrical source generator and the recording electrode, and applies when recording electrodes are not in direct contact with the source generator.

Each EEG electrode reflects the summated activity of cortical neurons in the underlying cortex approximately $6cm^2$ closest to the electrode. Therefore the distance between electrodes limits the spatial resolution of the EEG, on the one hand electrodes too far may not detect small areas of cortical epileptiform activity, however electrodes too close may lead to equality of potential and recording a straight isoelectric trace. Additionally, activity from deep structures such as basal ganglia and hippocampi are difficult to detect as voltage fields fall off at a square of the distance.

4.2 EEG features

4.2.1 The normal EEG and its variants

It is possible to identify a wide range of individual variability in the normal EEG. The human EEG usually shows activity in the range of 1 to 30Hz, with amplitudes in the range 20 to 300µV. The on going dominant electrical activity on the EEG is known as the background and is typically divided into four bands clinically; the delta band at 0.5 to 3Hz, theta band at 4 to 7Hz, alpha band at 8 to 13Hz and beta band at 14 to 30Hz. Table 4.1 summarises the major features of these EEG frequency bands. These features apply to adults and can be applied to children although children have some distinct EEG differences.

Table 4.1. EEG frequency bands and their features.

Band	Frequency range	Features
Alpha	8 to 13Hz	Dominant background rhythm. Highest amplitude in posterior regions. Usually symmetrical, may have slightly higher amplitude on the right. Difference in amplitude >50% suggest an identifiable abnormality Attenuated by eye opening and mental activity. Enhanced by eye closure and relaxation.
Beta	14 to 30Hz	Presence is normal, amount variable. Most evident frontally, symmetrical distribution. Low amplitude waves <20µV Increased by benzodiazepines and barbiturates.
Theta	4 to 7Hz	Common in young children and during drowsiness in adults. Not usually a well developed nor regular rhythm in healthy people
Delta	0.5 to 3Hz	Found during slow wave sleep. May be indicative of focal lesion or metabolic encephalopathy.

The EEG of a child is constantly changing due to maturation. The most striking difference in a child's EEG compared to an adult is the lower background frequencies and higher amplitude of discharges. The dominant background rhythm is dependent on the child's age (see Table 4.2) and evolves to adult levels within the second decade. Specific normal variants in childhood exist both in the awake and sleep EEG (marked by * in Table 4.3).

Table 4.2. Age related background frequencies.

Age	Background frequency
1 year	<5Hz
4 years	<6Hz
5 years	<7Hz
>8 years	<8Hz

In sleep the EEG is markedly changed. Specific sleep features are seen as the individual fluctuates between the three stages of non rapid eye movement sleep N1, N2 and N3 sleep and then rapid eye movement tonic and phasic sleep. Table 4.3 summarises these sleep features and stages based on the American Academy of Sleep Medicine 2007 manual[119]. The close relationship between sleep and epilepsy is undisputed. Non rapid eye movement sleep leads to a state of EEG synchronisation and promotes seizure propagation and activation of interictal epileptic discharges. Gibbs *et al.* first observed an increase of interictal epileptic discharges from 36% in awake EEGs to 82% with sleep in 500 epilepsy patients in 1947[120]. Since

then, several studies have observed an increase in interictal epileptic discharges in sleep in specific focal and generalised epilepsies[121,122]. Therefore, research participants were encouraged to sleep in order to improve the yield of capturing paroxysmal EEG discharges.

Table 4.3. Sleep features and sleep stages.

Discharges or sleep stage	Features
Hypnagogic hypersynchrony*	Discrete burst of diffuse high voltage rhythmic theta during drowsiness. Seen in early childhood.
Hypnopompic hypersynchrony*	Discrete burst of diffuse high voltage rhythmic theta during arousal. Seen in early childhood.
Vertex sharp waves	Principal component usually sharply contoured electronegative wave with maximal amplitude at Cz electrode. Seen in deeper stages of drowsiness, stage 1 and 2 of sleep.
K complexes	Biphasic wave, initial brief wave with subsequent slower wave superimposed with spindles. Appears in stage 2 sleep.
Sleep spindles	Identifies stage 2 of sleep. Frontal central waves, bilateral and synchronous with a frequency of 11 to 15Hz with an amplitude up to 30µV.
Benign epiletiform transients of sleep	Sharp, brief (~50ms) low amplitude (~50µV) waves that occur unilaterally or bilaterally during sleep especially in the temporal and frontal regions.
Positive occipital sharp transients of sleep	Monophasic, electropositive, sharply contoured occurring in bilateral occipital regions singly or in 4-5 second sequences. Most evident in stage 2 and 1 of sleep
N1 NREM 1 sleep	Shows attenuation of the alpha rhythm, slow rolling eye movements and the presence of vertex sharp waves.
N2 NREM 2 sleep	Presences of sleep spindles and K complexes.
N3 NREM 3 sleep	Delta waves occupy 20% to 50% of the recording.
R REM sleep	Delta waves occupy more than 50% of the recording.

*seen in childhood

4.2.2 Paroxysmal EEG discharges

Paroxysmal EEG discharges can be separated into potentially epileptiform discharges and benign EEG variants. Table 4.4 list the features of potential epileptiform discharges. Generalised spike and wave (GSW) can be considered the prototype epilepsy associated EEG phenomenon. Different types of GSW are commonly described. In MAE, an irregular GSW often preceded by polyspike is seen; this has quite a different appearance to 3Hz spike and wave complexes seen in childhood absence epilepsy. However as the main purpose of this

study was to measure the presence of epileptiform activity, all types of spike and wave complexes were grouped. Whilst most neurophysiologists would concur over epileptiform EEG discharges, benign variants are frequently debated.

Table 4.4. Paroxysmal potentially epileptiform EEG discharges.

Discharge	Feature
Spike	A pointed peak and a duration of 40 to 80ms (measured at mid-point between the baseline and the peak of the spike) easily differentiated from the background. Amplitude can be variable.
Sharp wave	A pointed peak between 80 and 200ms in duration (measured at mid-point between the baseline and the peak of the spike) easily differentiated from the background. The rising phase of a sharp wave is of the same order of magnitude as a spike but the descending phase is prolonged.
Spike and wave complexes	Spikes followed by a slow wave that occur repetitively in series.
Polyspikes	Multiple spike complexes with two or more spikes.

Benign variants are considered benign as they can occur in healthy individuals without epilepsy. However, these patterns merit careful consideration not only to avoid misinterpretation but also as they may yet represent undiscovered patterns of clinical significance. They can be age-associated features and are listed in Table 4.5.

Table 4.5. Benign EEG variants.

Discharge	Feature
Posterior slow waves of youth*	Waking, sinusoidal 2.5 to 4.5Hz slow wave that interrupts the background with voltage similar to the alpha voltage. Seen in later childhood and adolescence.
Rhythmic mid temporal discharges	5 to 7Hz monomorphic rhythm in burst or trains. Often sharply contoured and in mid-anterior temporal regions.
6/14 Hz positive spikes	Rhythmic arched waves with a smooth negative component and a spike like positive component. Occur singly or in bursts of 6Hz or 14Hz. Posterior temporal areas, during drowsiness and sleep.
Subclinical rhythmic EEG discharge of adults	Sharply contoured monophasic rhythmic delta or theta waves over temporo-parietal derivations. Bilateral or unilateral. More frequently seen in drowsiness and in the elderly.
Lamda waves	Sharp biphasic or monophasic waveforms occurring in occipital regions. Saccadic eye movements when presented with visual stimuli can evoke them.
Wicket spikes	Single spike like wave or intermittent trains of arc like monophasic waves at 6 to 11Hz. Usually in temporal regions.

4.2.3 Photoparoxysmal responses

The photoparoxysmal response (PPR) is defined as the occurrence of spikes or spikes and waves discharges in response to intermittent photic stimulation. Continuous flashes presented at 18Hz with eye closure are most likely to elicit a discharge in those who are susceptible. Sleep deprivation and/or awakening form a sleep stage can also increase the likelihood. PPR needs to be distinguished from photic driving which are rhythmic potentials in the occipital leads that are time locked with a photic stimulus and are considered normal. Waltz *et al.* distinguished the PPR into four types (see Table 4.6)[65].

Table 4.6. Classification of photoparoxsymal response.

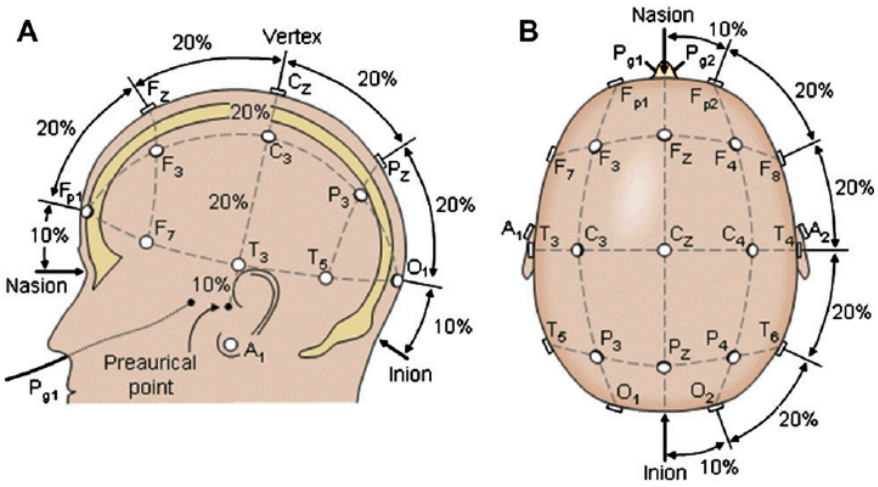
Type I	Occipital spikes
Type II	Local parietal-occipital spikes and biphasic slow waves spreading to frontal regions
Type III	Parietal-occipital spikes and biphasic slow waves spreading to frontal regions
Type IV	Generalised spikes or polyspike and wave

4.3 Recording conditions of research EEGs

Participants were invited to have EEG at one of three sites depending on participant preference, age of participant and availability of research appointments. The digital EEG systems used were: NicoletOne (Natus) at King's College Hospital, Neurofax EEG-1200 (Nihon Koden) at Evelina Children's Hospital and Xltek EEG32U amplifier (Natus) at St Thomas' Hospital. EEGs were administered and recorded by qualified clinical neurophysiology technicians.

The head was measured using landmarks based on the International 10-20 system (see Figure 4.2). The electrode position was prepped using Nuprep (Weaver and company, CO) to reduce scalp impedence. Silver-Silver chloride EEG electrodes are positioned onto the prepped with 10-20 EEG paste. The electrodes were placed based on the International 10-20 system. Ground and reference electrodes were placed between Pz and Cz and Cz and Fz respectively except in the Nihon Koden recordings where a linked C3-C4 reference was utilised. A lead I electrocardiogram (ECG) was placed on the bilateral shoulders and is used to identify pulse artefact and ECG breakthrough. EMG electrodes were placed on bilateral deltoids to capture myoclonic jerks and other movement activities. The EEGs were recorded digitally in a referential montage or bipolar montage.

Figure 4.2 The international 10-20 system EEG electrode placement.



Frequency filters were set to reduce the size of non-relevant and artefactual frequencies. High frequency filter was set at 70Hz and low frequency filters at 0.5Hz. The band pass filter was set at 50Hz to eliminate common electrical interference from UK mains power lines. Impendences did not exceed 5kΩ. The Nyquist frequency represents the minimum rate at which a signal can be sampled without aliasing of the signal. This is twice the highest frequency present in the signal we wanted to record and therefore was set above 200Hz (see individual sampling frequency in Table 5.1). Participants were instructed to have a period of relaxed awake eye closure recording before encouraging natural sleep. The total recording time was a minimum of 30 minutes.

4.4 Activation procedures

In order to increase the yield of capturing epileptiform activities, activation procedures of hyperventilation and photic stimulation were performed and subjects were asked to sleep.

Hyperventilation

Hyperventilation creates vascular hypocarbia and in turn cerebral vasoconstriction. The subject is asked to take deep breaths and hyperventilate for 3 minutes. In the child, the subject is asked to blow on a windmill. Hyperventilation is contraindicated in persons with severe cardiac or pulmonary disease, acute or recent stroke, significant large vessel cerebrovascular disease and sickle cell anaemia. This technique causes a slowing of background frequencies consisting of rhythmical diffuse high amplitude synchronised delta slow waves. This is considered a normal phenomena if there is resolution of slowing within one minute after hyperventilation

ceases. Hyperventilation is useful in eliciting absence seizures with 3Hz GSW discharges, and focal slowing and sharp waves.

Photic stimulation

Photic stimulation is commonly used to elicit PPR. It commonly causes photic driving, a normal rhythmic potential in the occipital leads time locked to the frequency of the light stimulus. A light source is placed approximately 30cm from the patient. Intermittent photic stimulation was performed at 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30 and 50Hz.

Sleep

EEGs were carried out in a quiet and isolated room. Subjects were asked to lie on a coach or bed in the dark with their eyes closed, and encouraged to sleep. A period/s of drowsiness in between awake and sleep stages throughout the recording is frequently seen.

4.5 Qualitative analysis of EEG features

The entire EEG recording was visually inspected independently and then re-reviewed by a Consultant Clinical Neurophysiologist (Dr Sushma Goyal, Professor Michalis Koutroumanidis). Atypical EEG features were reviewed with at least two different montages.

The following features were documented on a proforma (see appendix D):

1) Background rhythm: frequency, responsiveness to eye closure and symmetry.
2) Presence, frequency and distribution of spikes, sharp waves and generalised spike wave complexes.
3) Response to hyperventilation.
4) Response to photic stimulation and classification (criteria in Table 4.6).
5) Sleep stage (criteria in Table 4.3).
6) Presence of benign variants (see Table 4.5) or atypical EEG patterns.

4.6 Population prevalence of EEG discharges

EEG discharges in healthy individuals have been investigated in several different settings. The first adult study was published in 1953[123]. The authors investigated 682 air force applicants aged 17 to 24 years with an eight channel EEG, at rest, with hyperventilation and photic stimulation, and with Metrazol, a GABA antagonist. Whilst the subjects were considered normal, 36 out of 682 were later described to have definite neurological findings. The authors

also added that subsequently 203 subjects were rejected from service due to a combination of personality disturbances, neurological disturbance on examination or history, medical or miscellaneous reasons. EEG paroxysmal discharges were present in 2.6% at rest, 2.9% with hyperventilation and 5.1% with hyperventilation and photic stimulation prior to metrazol administration. The incorporation of these unfiltered subjects may explain why subsequent studies have reported a lower prevalence of epileptiform activity. Table 4.7 summaries EEG studies in healthy adults.

Table 4.7 Summary of EEG studies in healthy adults.

Study	Recording conditions	Subjects	Cases with paroxysmal discharges
Buchthal & Lennox 1953[123]	Eight channel EEG At rest, HV, photic	682M 17 to 24 years	2.6% (rest) 2.9% (HV) 5.1% (HV and photic)
Gregory *et al.* 1993[124]	Eight channel EEG At rest, HV, photic	13658M 17 to 25 years	0.5% (69M) GSW or Focal 21 cases PPR 44 cases At rest and PPR 4 cases
Jabbari *et al.* 2000[125]	16 and 18 channel At rest , HV, photic, sleep (n=96)	100M 18 to 45 years	0% RMTD 1 case

HV hyperventilation, photic photic stimulation, M male, GSW generalised spike wave, PPR photoparoxysmal response, RMTD rhythmic mid temporal discharges

The prevalence of epileptiform discharges in children range from 0.76% to 13.2% across different studies. This broad range may be due to a number of reasons: differences in recording conditions, selection of control individuals, age dependent EEG features, inconsistent classification of EEG features recorded and challenges in interpretation of the paediatric EEG. Table 4.8 summarises the key features of these EEG studies in healthy children. It is worth noting that Eeg Olofsson *et al.* recorded paroxysmal discharges resulting in an overall prevalence of 13.2%[126]. This could have included non-epileptic discharges; such as photic driving within the PPR group and bilateral synchronous paroxysmal activity during drowsiness which may represent hypnogogic hypersynchrony as they were predominantly in cases under 6 years[126]. Additionally Borusiak *et al*. and Bihege *et al*. had cohorts of children with minor head trauma, it is unclear how this might have affected EEG features[127,128].

Table 4.8. Summary of EEG studies in healthy children.

Study	Recording conditions	Subjects	Cases with paroxysmal discharges
Eeg-Olofsson *et al.*[126] 1970	Paper EEG At rest, HV (82.6%), photic (81.4%), sleep (80.5%)	743 (284M, 389F) 1 to 15 years	13.2%* Focal 2.4% GSW 0.3% PPR 8.3% Sleep 8.1%
Cavazzuti *et al.*[129] 1980	Paper EEG At rest, HV	3726 (1988M, 1738F) 6 to 13 years	131 (88M, 43F) (3.54%) GSW 41 cases Focal 52 cases Multifocal 11 cases CTS 27 cases
Okubo *et al.*[130] 1994	Paper EEG 8 electrodes At rest, HV	1057 (563M, 494F) 6 to 12 years	53 (33M, 20F) (5%) GSW 10 cases Focal 3 cases Multifocal 2 cases CTS 37 cases
Borusiak *et al.*[127] 2010	Digital EEG, 21 electrodes At rest, HV, photic, sleep (15.7%)	382 (226M, 156F) 6 to 13 years minor head trauma	25 (11M, 14F) (6.5%) GSW 4 cases Focal 4 cases Multifocal 9 cases CTS 7 cases
Bihege *et al.*[128] 2015	Digital EEG 10-20 system At rest, HV, photic, sleep in some	393 (197M, 196F) 1 to 5 years minor head trauma	3M (0.76%) GSW 1 case Mutlifocal 1 case CTS in 1 case

HV hyperventilation, photic photic stimulation, M male, F female, GSW generalised spike wave, PPR photoparoxysmal responses, CTS centro temporal spikes, *18 cases had more than one feature.

4.7 Genetic basis of the EEG

Evidence for the genetic basis of the EEG exists across a range of EEG features and can be gleaned through family studies, linkage analysis studies and quantitative EEG studies. The familial occurrence of non-epileptiform discharges was recognised in families when Koshino *et al.* reported the incidence of the mμ rhythm in 13/35 relatives of 14 families[131]. Specific epileptiform features have also been investigated. Centrotemporal sharp waves in rolandic epilepsy families were found to segregate in an autosomal dominant mode of inheritance[132]. GSW discharges were seen in 32/83 (38.5%) of siblings of 54 probands with symptomatic GTCS, and only seen in sleep in 15 siblings[133]. Doose investigated specific EEG features of PPR, GSW or dysrhythmias with his initial description of MAE and identified abnormalities in 35.4% of 72 siblings and 79 parents[5] (see section 1.7 for further details). PPR were identified to be heritable by Waltz *et al*. when he demonstrated that type IV PPR occurred more often in probands with epilepsy and their siblings than respective controls[65]. PPR deserves further

mention as it represents an example of an extension to exploration of an aetiological genetic locus through linkage analysis.

Genome wide linkage studies have identified several different linkage regions associated with PPR. Pinto *et al*. mapped two susceptibility loci 7q32 (HLOD 3.47) and 16p13 (HLOD 2.44) in 16 PPR (type II to IV) myoclonic epilepsy families[66]. Tauer *et al.* investigated two different phenotypic family subgroups with either pure PPR (n=19) or PPR/IGE (n=25). They identified locus on 6p21.2 to predispose to PPR itself whereas locus on 13q31.1 to confers susceptibility to IGE[67].

The heritability of the EEG can also be traced with EEG power spectral studies. By studying 213 twin pairs, spectral power for the four major frequency bands; alpha, beta, theta and delta demonstrated average heritabilities of 89%, 86%, 89% and 76% respectively[134]. The next section will focus on quantitative EEG and power spectral analysis. Although not one of the original aims of the project, power spectral analysis was carried out in order to explore the relationship of background rhythms in subjects.

4.8 Quantitative EEG

Quantitative EEG is the mathematical processing of transforming digitally recorded EEG data into numerical data to derive quantitative patterns that may correspond to diagnostic information or cognitive effects. One of the parameters of quantitative EEG analysis includes Fourier analysis, which converts the original EEG data into a voltage by frequency spectral graph, commonly known as the power spectrum.

4.8.1 Power spectral analysis

Power spectral analysis was performed comparing the EEGs in groups of fathers, mothers and siblings. Power spectral analysis involves decomposing the EEG mathematically into a number of pure sinusoidal components, each of a different frequency, which when added together yield the original signal. The magnitude corresponds to the amount of energy that the original EEG possesses at each frequency. The conversion of signal from the time domain into the frequency domain (and vice versa) is known as Fourier transform.

The EEG segments per individual was selected as follows. The earliest two 10 second epochs were chosen based on the following criteria; (1) eye closure wakefulness with the presence of

alpha activity maximal in the posterior leads, (2) the absence of artefacts (eye blinks, eye movements, muscle contractions, movement of recording electrodes) and epileptiform activity, (3) the absence of patterns indicating drowsiness or sleep and (4) not during activation procedures or within 3 minutes of hyperventilation. Epoch selection was done independently and then jointly with a qualified clinical neurophysiological technician (Stewart Smith).

All power spectral analysis was carried out using FieldTrip[135], a MATLAB-toolbox for the analysis of EEG and other electrophysiology data, which is freely available from http://www/ru.nl/neuroimaging/fieldtrip. Custom scripts were implemented in Matlab R2016a (Mathworks, Natick, Massachusetts, USA). Custom MATLAB scripts utilising Fieldtrip functions for EEG data preprocessing and analyses were written originally by Dr Adam Pawley and Dr Chayanin Tangwiriyasakil, and adapted by Dr Chayanin Tangwiriyasakil for this analysis. The EEG signals were first re-sampled individually to a uniform 256Hz sampling frequency. They were then de-trended to remove signal drifts, and bandpass filtered between 0.5 to 70Hz using a 4^{th} order butterworth filter. The power spectral density was then estimated using Welch's technique (2 second long window with 50% overlap) using a Hanning window.

Power for five frequency bands (delta 1 to 3 Hz, theta 4 to 7 Hz, alpha 8 to 14 Hz, beta 15 to 30 Hz, and gamma 30 to 70Hz) was estimated by numerically integrating the power spectral density over the band and divided by a total power of all five bands. To avoid possible amplitude differences in recorded EEG signals across subjects, due to variation caused by different anatomic features such as size and shape of the cranium; a normalized power of each band was estimated by dividing the power of each frequency band with its total power. The total power was estimated by numerically integrating the power spectral density over 1 to 70Hz.

The Kruskal-Wallis test by ranks was used to test if the average normalized powers at each frequency band were significantly different between the three groups (fathers, mothers, siblings). The Kruskal Wallis test also called the one-way analysis of variance (ANOVA) on ranks, is a rank based non parametric test used to determine statistical differences of an independent variable between two or more groups of equal or different sizes. Wilcoxon rank's test was then applied to compare mean normalized power between the three groups. This comprised a total of 15 comparisons: fathers and mothers, mothers and siblings, fathers and siblings multiplied by the five different frequency bands. The test was re-applied with a 5%

false discovery rate (type 1 error) correction to account for the 15 comparisons. A *P* value ≤0.05 was considered significant.

Chapter 5 Result of EEG studies

5.1 EEG research participants

Thirteen families with 38 first-degree family members were recruited to the EEG arm of the study (see Table 5.1). This comprised of 12 fathers (mean age 47.2 years, SD 7.13), 13 mothers (mean age 43.7 years, SD 6.12), 8 brothers (mean age 12.7 years, SD 4.84) and 5 sisters (mean age 11.6 years, SD 7.56). Family size ranged from three to six individuals. Four families had incomplete studies. Family 506, 518 and 526 were incompletely recruited with one missing family member in family 506 and 526 and two missing family members in family 518. Family 602 had a child with Down's syndrome who was unable to carry out the EEG. Some participants (±) did not receive activation procedures of hyperventilation and photic stimulation as the wrong EEG recording protocol was completed for the study in error. EEGs were performed at all three sites; St Thomas' Hospital (n=17), King's College Hospital (n=11) and Evelina Children's Hospital (n=10).

Table 5.1 Details of EEG participants

Family ID	Subject ID	Relation	Age (years)	EEG site	EEG duration (hours:minutes:seconds)	Sampling frequency (Hz)
3002	3002201	Father	54y	STH	1:09:04	256
	3002202±	Mother	50y	KCH	48:50	256
	3002302±	Sister	19y	KCH	57:38	256
	3002303	Brother	17y	STH	55:19	256
	3002304	Brother	17y	STH	49:14	256
505	10505	Father	45y	ECH	45:24	500
	20505	Mother	40y	ECH	54:19	500
	30505	Brother	8y	ECH	1:27:53	500
	40505	Sister	2y	ECH	54:25	500
506*	10506	Father	44y	STH	45:02	256
	20506	Mother	44y	STH	1:10:46	256
518*	10518	Father	50y	ECH	44:53	500
	20518	Mother	40y	ECH	36:43	500
524	10524	Father	46y	ECH	43:41	500
	20524	Mother	50y	ECH	37:31	500
	30524	Brother	9y	ECH	43:41	500
526*	20526±	Mother	-	KCH	1:05:15	256
	30526±	Brother	12y	KCH	55:02	256
528	10528	Father	59y	STH	56:54	256
	20528	Mother	47y	STH	58:10	512
	30528	Sister	16y	STH	45:14	256
539	10539	Father	41y	STH	47:52	256
	20539	Mother	44y	STH	49:42	256
554	10554	Father	56y	STH	54:59	512
	20554	Mother	52y	STH	49:51	256
	30554	Brother	20y	STH	35:51	256
560	10560±	Father	34y	KCH	56:39	256
	20560±	Mother	35y	KCH	58:36	256
571	10571	Father	40y	STH	1:06:20	512
	20571	Mother	33y	STH	53:58	512
	30571	Brother	5y	ECH	1:00:23	500
574	10574±	Father	48y	KCH	58:48	256
	20574+	Mother	41y	KCH	59:53	256
	30574±	Sister	5y	KCH	58:11	256
602*	10602	Father	50y	STH	55:28	256
	20602	Mother	49y	STH	53:41	256
	30602±	Sister	16y	KCH	1:02:34	256
	40602±	Brother	14y	KCH	59:25	256

* Incomplete families, ± Hyperventilation and photic stimulation not performed, y years, STH St Thomas' Hospital, KCH King's College Hospital, ECH Evelina Children's Hospital.

5.2 Qualitative EEG visual analysis results

All subjects had symmetrical posterior dominant rhythms within the alpha range. A period/s of drowsiness in between awake and sleep stages throughout the recording was frequently seen. All the subjects had evidence NREM1 and 2 sleep on their EEG apart from subject 10506 and subject 20602. The mean duration of the EEG records was 54.01 minutes (SD 10.78).

5.2.1 Epileptiform EEG features

Two subjects had epileptiform features, both were asymptomatic and had no history of seizures at the time of EEG recording. The two subjects were subject 20526 (mother) with epileptiform left frontal temporal sharp waves (see Figure 5.1) and subject 30574, a 5-year-old sister with 3Hz generalised spike wave (see Figure 5.2 and the same epoch with lower sensitivity in Figure 5.3). Subject 30574 subsequently developed absence seizures and a repeat clinical EEG with activation procedures confirmed a diagnosis of childhood absence epilepsy.

Figure 5.1. EEG demonstrating left frontal temporal sharp waves in subject 20526.

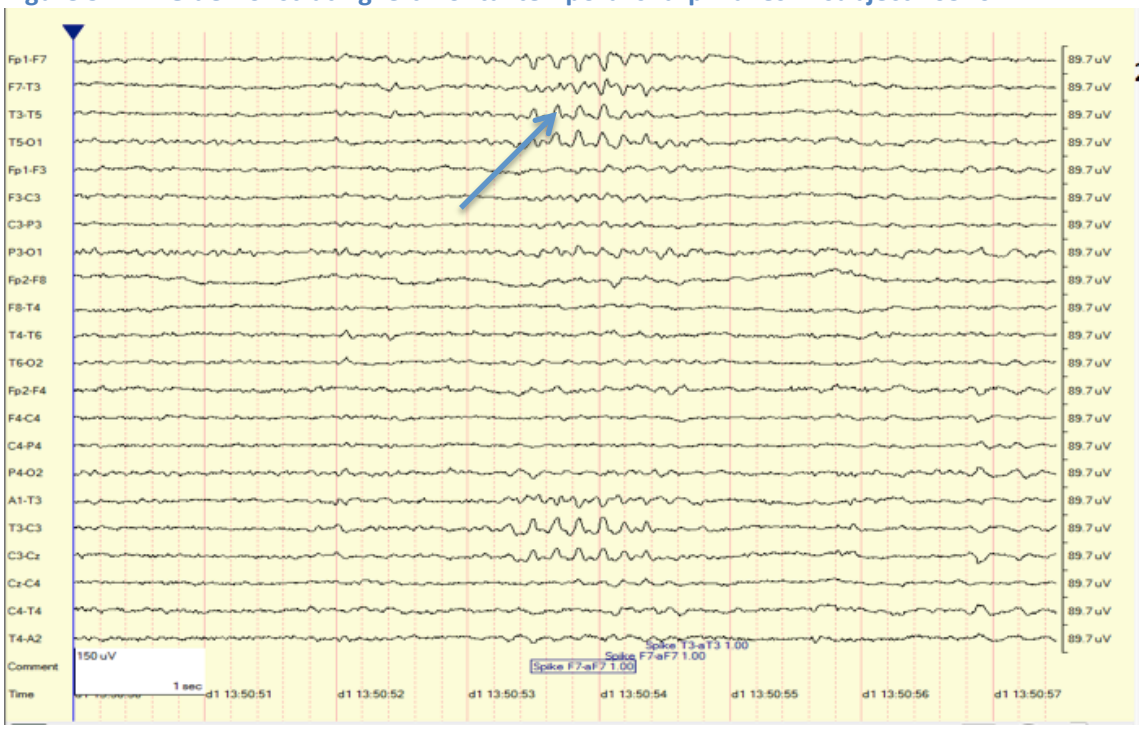


Figure 5.2. EEG demonstrating 3Hz generalised spike wave in subject 30574.

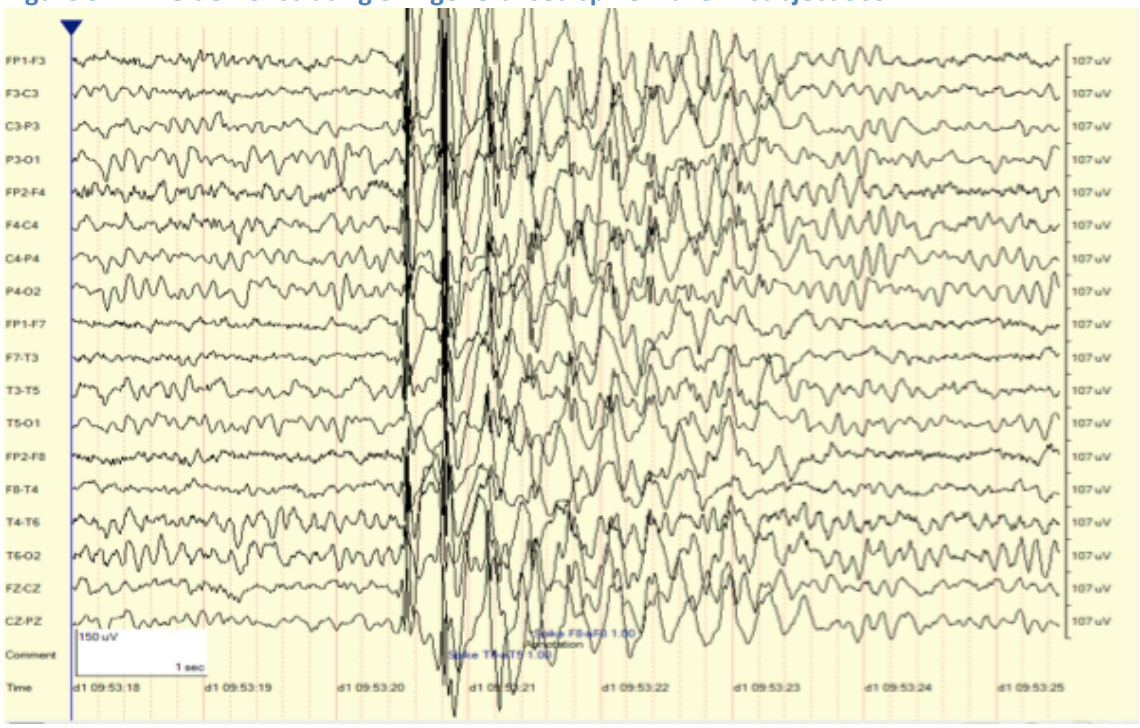
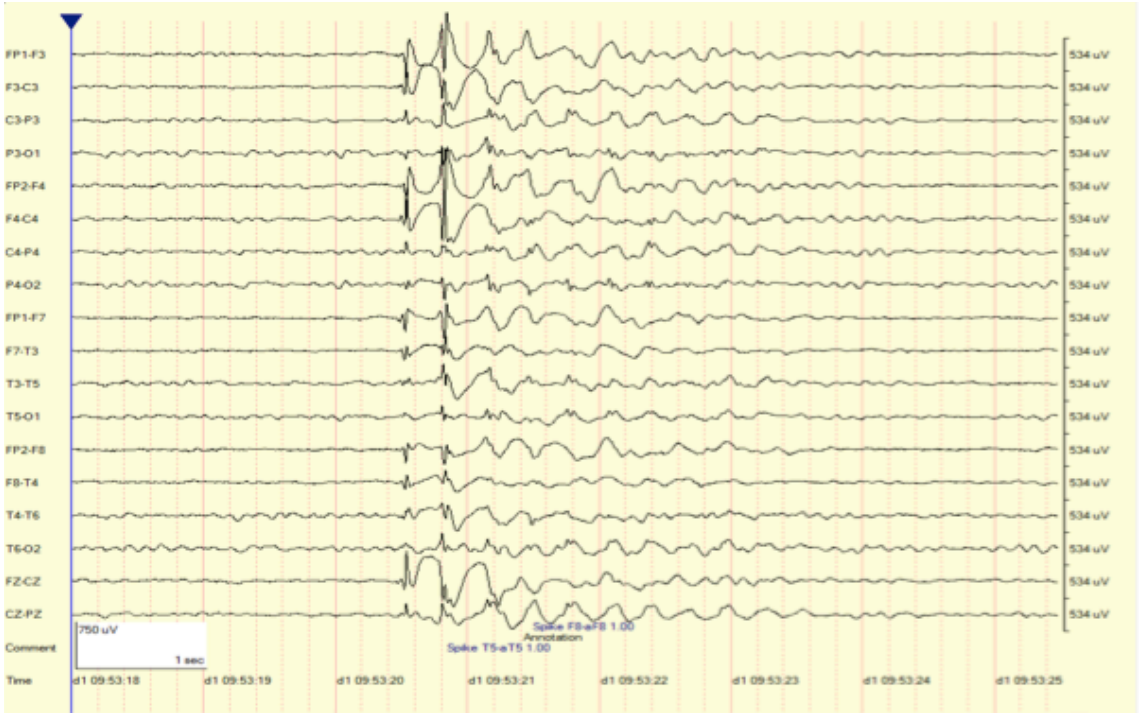


Figure 5.3 Same epoch as figure 5.2 in subject 30574 with lower sensitivity.



This epoch with lower sensitivity demonstrates the morphology and duration of the spike wave discharges.

PPR, the most frequently seen abnormal finding in Doose's MAE family EEG study, was not recorded. Table 5.2 shows the comparison of epileptiform EEG features with Doose *et al.*[5] using Fisher's exact test and the relative risk (RR) of recording each EEG feature in Doose *et al.*'s cohort compared to this cohort. This shows that there was no statistical difference in the identification of specific traits in both cohorts. However, the RR of observing PPR (siblings= 5.61, parents=3.71) was much higher in Doose *et al.*'s cohort compared to this MAE cohort. Several possible reasons could account for this.

Table 5.2. Comparison of epileptiform EEG features with Doose *et al.*'s study.

MAE family members	Epileptiform waveforms	Doose *et al.*[5]	MAE cohort	RR Doose *et al.* / MAE cohort	*P* value
Siblings	PPR	20/72 (27.8%)	0/9 (0%)	5.61	0.21
	GSW or Sharp wave	6/72 (8.33%)	1/13 (7.6%)	1.08	0.93
Parents	PPR	6/69 (8.69%)	0/19 (0%)	3.71	0.36
	GSW or Sharp waves	1/69 (1.45%)	1/25 (4%)	0.36	0.46

Two-tailed *P* value determined using Fisher's exact test.
RR relative risk, PPR photoparoxysmal reaction, GSW generalised spike wave

First, recording conditions were vastly different. Doose carried out paper EEGs with 8 channels whereas all EEGs here were performed on digital recordings with electrodes laid out in the 10-20 systems. The lack of activation procedures, specifically photic stimulation in some of this study's subjects would have compromised results. The small sample size in this MAE cohort and incomplete families recruited limits the power of discovery. The definitions of PPR and EEG abnormalities may have differed. In addition, Doose recorded dysrhythmias, which were not accounted for in this study group. An attempt is made to address this with quantitative EEG analysis (see section 5.3).

Comparison with population prevalence of epileptiform EEG features is shown in Table 5.3. The MAE parents and siblings were divided into age groups above and below 16 years to reflect differences in adult and paediatric EEG studies. Gregory *et al*. was used as the adult control study[124] and Borusiak *et al*. was used as the paediatric control study[127]; these controls were selected due to similar recording conditions. This shows that the presence of epileptiform discharges in adults (>16 years) was higher than controls (*P*=0.05, RR 6.82), but did not reach statistical significance in the siblings (<16 years) *(P*=0.58, RR 1.69).

Table 5.3. Comparison of epileptiform EEG features with controls.

Age groups	MAE cohort	Controls	RR MAE cohort / controls	*P* value
> 16 years	1/29 (3.4%)	69/13658 (0.5%)[124]	6.82	0.05
< 16 years	1/9 (11.1%)	25/382 (6.5%)[127]	1.69	0.58

P value calculated with chi squared with yates correction

5.2.2 Atypical EEG features

Table 5.4 provides a summary of the atypical patterns that were observed in the participants. These patterns were observed in mothers or fathers but not in siblings. Some patterns had features of well-described benign variants, e.g. rhythmic mid-temporal discharges and were classified as such. Other patterns are simply described based on their visual appearance. The significance of these patterns is discussed through literature review along with the proband characteristics.

Table 5.4: Summary of benign and atypical EEG features recorded

Subject ID	Description of pattern
20554	Rhythmic mid-temporal discharges
20602	Rhythmic mid-temporal discharges
20528	Post rhythmic theta activity during drowsiness
10539	Bilateral theta rhythms
20506	Left temporal slow activity
10528	Runs of slow activity over temporal regions, more on the left

Rhythmic mid-temporal discharges

Rhythmic mid-temporal discharges (RMTD) or psychomotor variant have EEG characteristics in adults consisting of a burst of rhythmic, sharply contoured waves with a frequency of 5 to 7Hz occurring over the temporal regions independently or on both sides. It is seen more frequently in the drowsy state. RMTD is reported to have an incidence of 0.1 to 0.3% with little sex preference[136,137]. Few conflicting studies exist for the clinical significance of RMTD. RMTD appeared to have a subclinical role in the occurrence of epileptic seizures[137] but was also reported in patients without epileptic seizures[138,139]. RMTD has also been associated in patients

with headache and psychiatric disorders[137] and a single case of a 9 year old child with an underlying structural lesion[140].

RMTD was recorded in two mothers 20554 and 20602. Both subjects had no relevant personal medical history. The relevant probands had similar onset of their epilepsy course but proband 00602 went into seizure remission whereas 00554 had ongoing seizures. Family 602 had a strong family history of epilepsy with a brother 10602 with epilepsy but there was no family history with family 554. No significance of the RMTD patterns could be inferred.

Figure 5.4. EEG demonstrating rhythmic mid-temporal discharges in subject 20602.

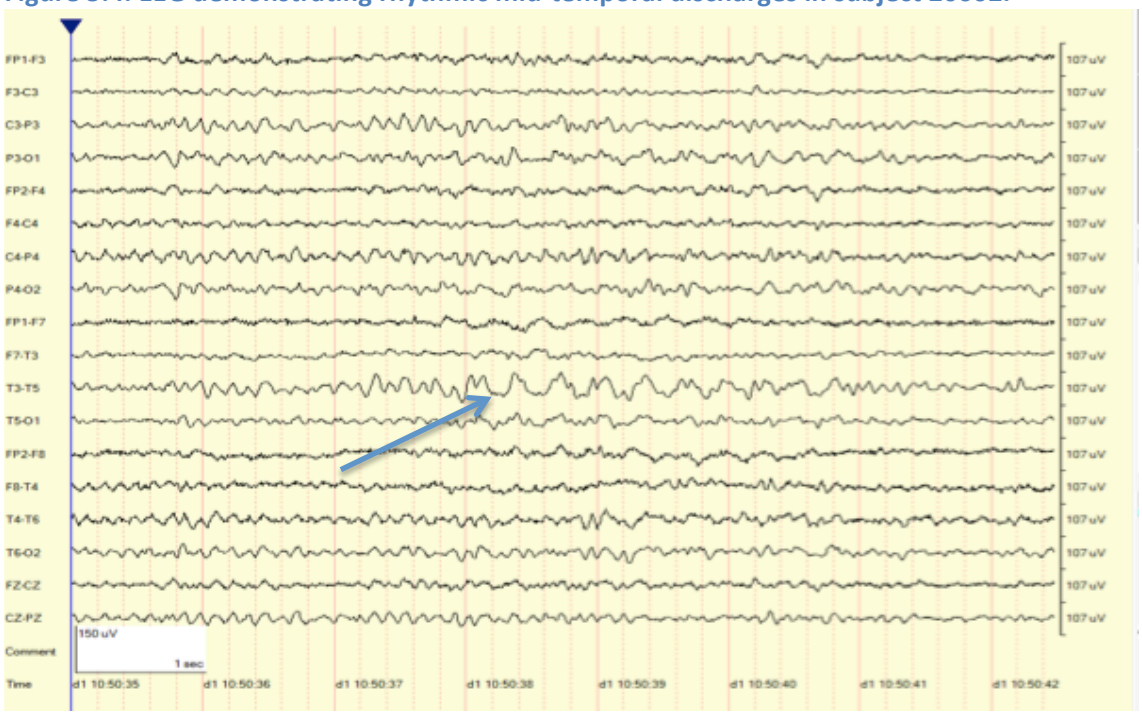
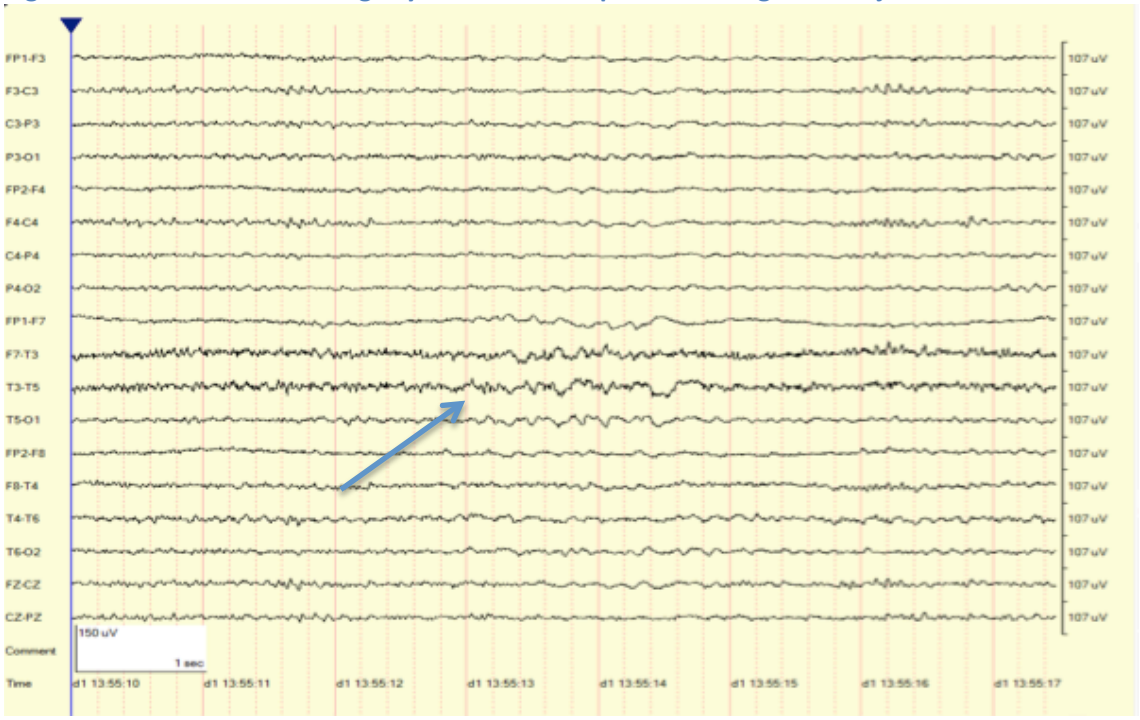


Figure 5.5: EEG demonstrating rhythmic mid-temporal discharges in subject 20554.



Theta rhythms

Theta rhythms are 4 to 7Hz and are seen more frequently in drowsiness and in young children given their slower dominant background rhythms. Doose *et al*. described different sub types of theta rhythms and recognised them as a genetic trait[68]. He observed that they are seen more frequently in young siblings of epilepsy probands and also acted as a risk factor with 63% of 'theta positive' children subsequently developing febrile convulsion[68]. Subsequently quantitative EEG studies in adolescent twin pairs demonstrated a heritability of 89% for theta power[134], a finding which was supported by Tye *et al*. who calculated a cross twin within trait correlation for theta power in monozygotic twins of r=0.79 compared to dizygotic twins of r=0.13[141].

Theta rhythms were recorded in two adult patients. Subject 10529 is a 41-year-old healthy father; he had no personal or family history of seizures or other neurological illnesses (see Figure 5.6). However subject 20528 (mother of subject 00528), although herself asymptomatic has a niece, (cousin of subject 00528) with childhood absence epilepsy (see Figure 5.7). This is potentially interesting as it supports the conjecture of theta rhythms as a genetic trait with a multifactorial role in epilepsy.

Figure 5.6. EEG showing bilateral theta rhythms in subject 10539.

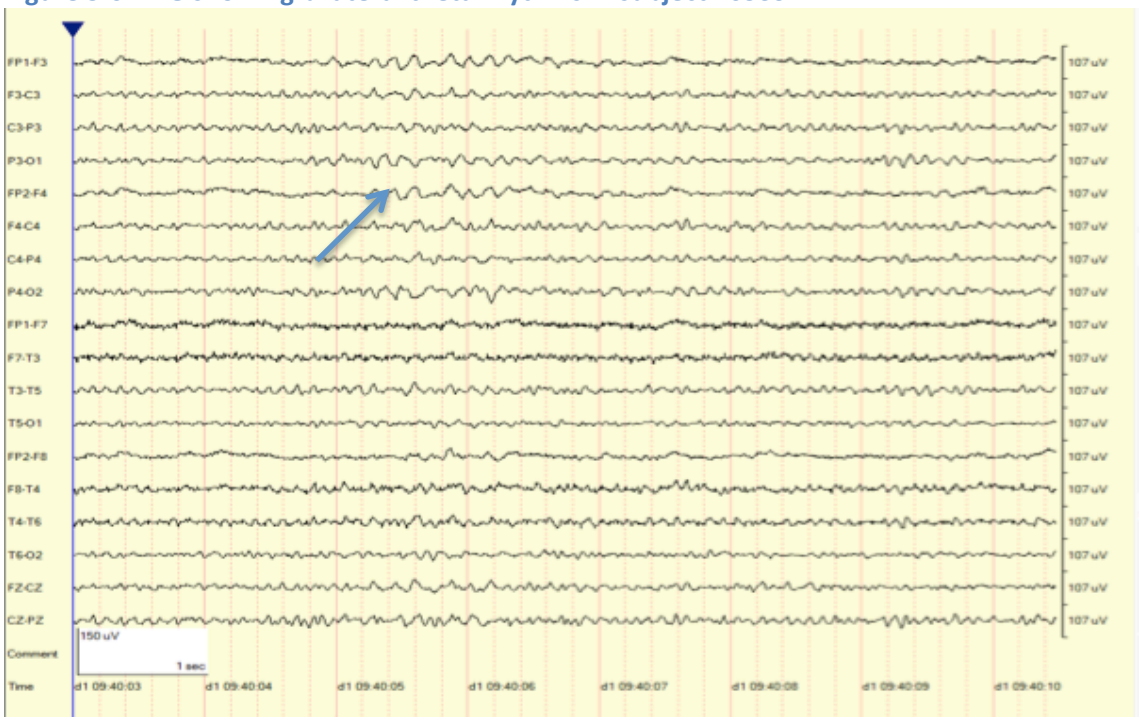
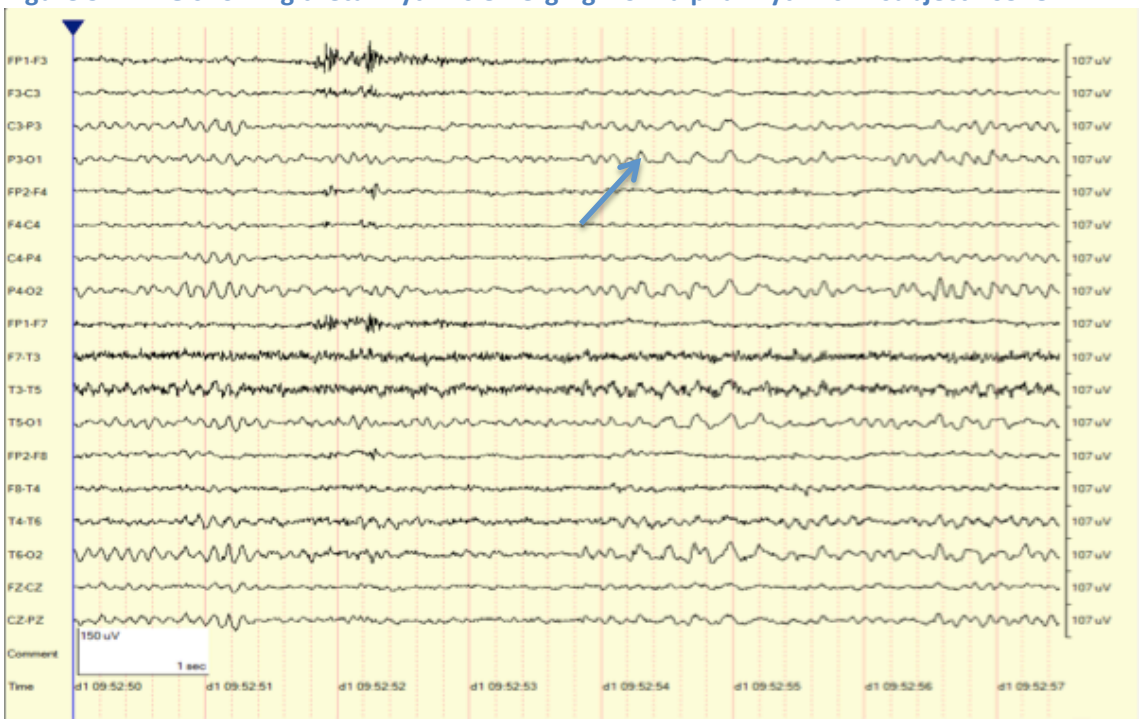


Figure 5.7. EEG showing theta rhythms emerging from alpha rhythms in subject 20528.



Other atypical patterns

Atypical patterns in the temporal regions of slow activity in subject 20506 (see Figure 5.8) and runs of slow activity in subject 10528 (see Figure 5.9) were also recorded. As an isolated finding in a healthy individual, it is difficult to speculate what these may represent. Intermittent focal slowing often reflects disordered intra-cortical connectivity. This can be transient (e.g migraine, post ictal, metabolic), structural (e.g. ischaemic stroke, tumour) or even represent an epileptogenic focus. The locations of these patterns, the temporal regions are the sites for many well-recognised benign variants (wicket spikes, RMTD, subclinical rhythmic EEG discharge of adults), and it is possible that these are atypical manifestation of these variants.

In order to further interpret all these patterns and their significance, quantitative EEG analysis in the form of power spectral analysis was carried out.

Figure 5.8. EEG showing left temporal slow activity in subject 20506.

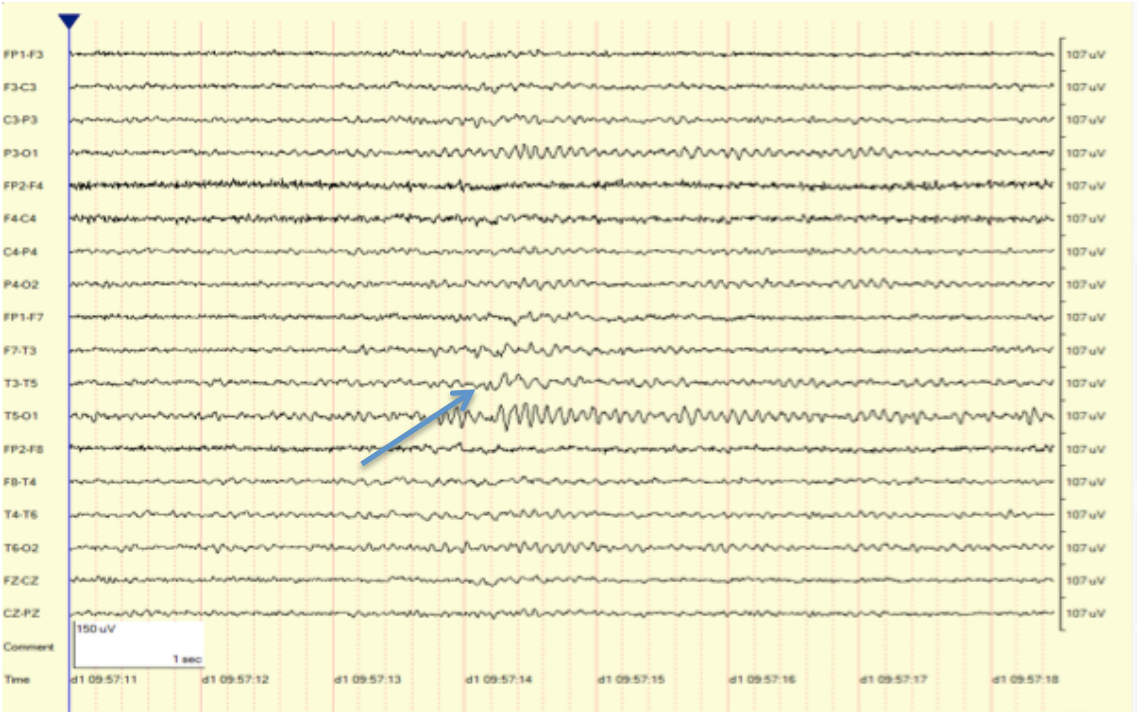
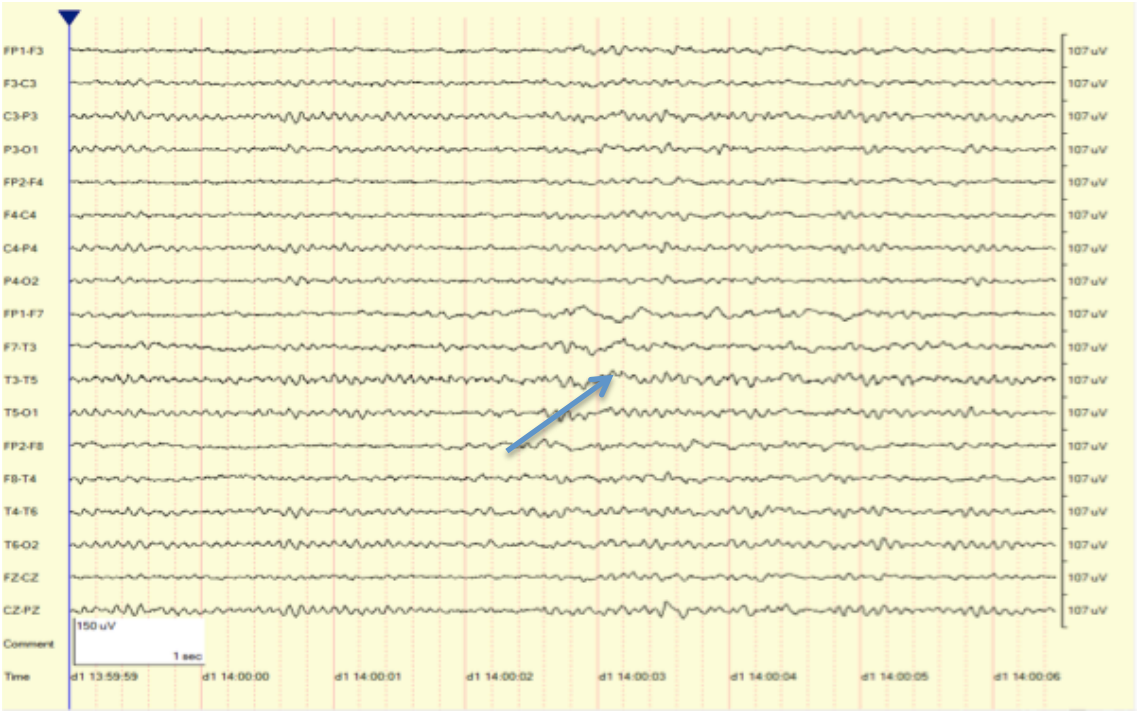


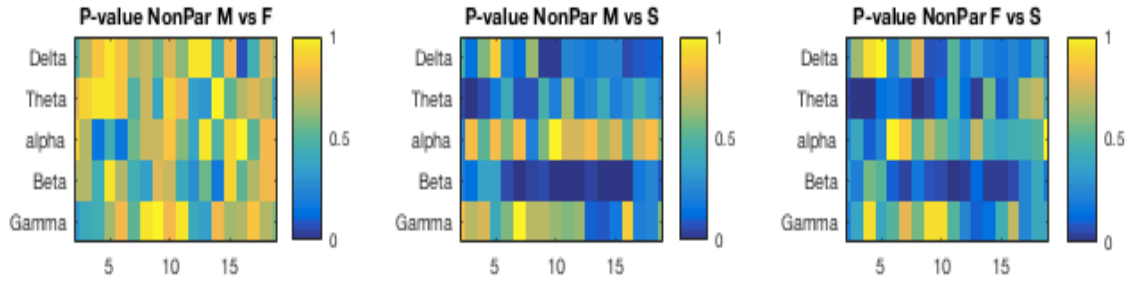
Figure 5.9. EEG showing runs of slow activity over left temporal regions in subject 10528.



5.3 Quantitative EEG power spectral results

Three groups comprising 13 mothers (M), 10 fathers (F) and 6 siblings (S) were analysed for power spectral in five frequency bands (delta 1 to 3Hz, theta 4 to 7Hz, alpha 8 to 14Hz, beta 15 to 30Hz, and gamma 30 to 70Hz). Four subjects (10602, 10571, 30426, 20574) were excluded as it was not possible to identify a suitable 10-second epoch that satisfied the selection criteria (see section 4.8.1). Figure 5.10 shows the Wilcoxon rank test comparing the power of the five frequency bands and the 19 channels in groups of mothers, fathers and siblings. Channel (ch) number representation were as follows: ch2=C3, ch3=C4, ch4=O1, ch5=O2, ch6=A1, ch7=A2, ch8=Cz, ch9=F3, ch10=F4, ch11=F7, ch12= F8, ch13=Fz, ch14=Fp1, ch15=Fp2, ch16=Fpz, ch17=P3, ch18=P4, ch19=Pz. Ch1 was removed due to an inconsistent/absent recording.

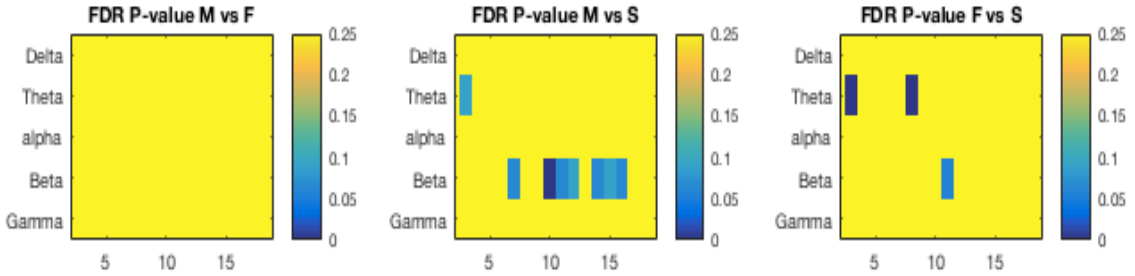
Figure 5.10. Wilcoxon rank test comparing mothers, fathers and siblings.



M Mother, F Father, S siblings. NonPar Non parametric. Each subplot shows comparisons of the five frequency bands (y axis) and 19 channels (x axis) and the *P* values obtained. *P* values close to zero are blue in colour whereas *P* values close to one are yellow.

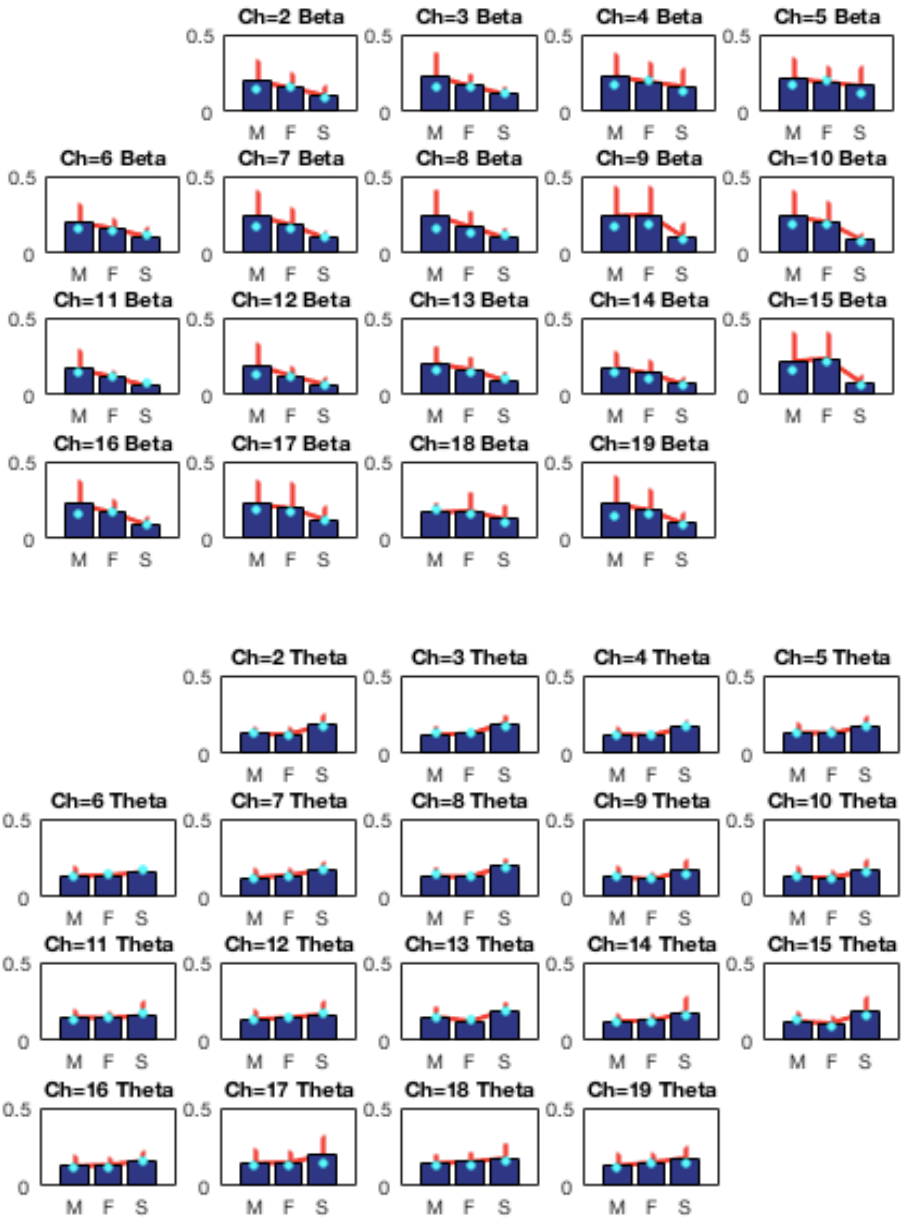
Figure 5.11 presents the results of these rank tests corrected with a false discovery rate (FDR) of 0.05. This demonstrated significant differences in the beta and theta bands, and the normalised power spectral of these bands are shown in Figure 5.12.

Figure 5.11 False discovery rate correction of Wilcoxon rank test.



FDR false discovery rate, M mothers, F fathers, S siblings. Blue colour denotes the significant difference between the mean of two groups (FDR *P* ≤0.05). Light blue to green colour denotes the borderline difference (0.05< FDR *P* <0.15), and yellow denotes no difference (FDR *P* >0.15).

Figure 5.12. Comparison between groups across channels of normalised power spectral in beta frequency band (above) and theta frequency band (below).



M mothers, F fathers, S siblings. Mean values are represented by the top margin of each blue column, red bars represent SD, and light blue dots represent median values.

Comparison of Mothers vs Fathers

There were no significant differences in the normalised power spectrum between mothers and fathers after FDR correction.

Comparison of Mothers vs Siblings

Significant differences were seen in the theta frequency band in ch3 (P=0.0125), and in the beta band on ch7 (P=0.0066), ch10 (P=0.0014), ch11 (P=0.0066), ch12 (P=0.0125), ch14 (P=0.0066), ch15 (P=0.0125) and ch15 (P=0.0066).

Comparison of Fathers vs Siblings

Significant differences were seen in the theta frequency band in ch3 (*P*=0.0011) and ch8 (*P*=0.0027); and in the beta frequency band in ch11 (*P*=0.0047).

The increased power spectrum in the theta frequency band in siblings compared to parents could be possibly explained by slower dominant background rhythms in children, although it is not possible to infer specific association of why this occurred in ch3 (C4) and ch8 (Cz). The increased beta power in mothers and fathers compared to siblings occurs mostly in the frontal and frontal polar leads, could be explained by a normal presentation of the beta rhythm more commonly seen in adults.

Comparisons with controls

Differences in the power spectrum within the theta and beta frequency bands can be explained by age dependent and physiological EEG inter variability. Hence, control data for adults (Dr Isabella Premoli) and children were collected (Dr Charlotte Tye). Unfortunately, due to differences in how the EEG data were formatted, this was not compatible with my matlab scripts and it was not possible to carry out further analysis.

Chapter 6 Methods: Genetics

6.1 DNA and RNA processing

Blood or saliva deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) was collected from the case, father, mother and other relevant family members affected with seizures and/or neurological disorders. Saliva DNA and RNA was the main mode of collection, blood was collected when subjects were directly recruited or where saliva collection was difficult.

Samples were processed and stored in two facilities. This was either by the Social Genetic Developmental Psychiatry (SGDP) Biobank and Bioresource or by the Pal Lab group at the Wohl Clinical Neuroscience Institute.

Blood DNA collection

Blood for DNA extraction was collected in a 5ml (child) or 10ml (adult) EDTA tube (Sarstedt, part number 03.1068.001) and blood for RNA extraction was collected in a 5ml Tempus tube (Thermo Fisher Scientific, Life Technologies, part number 4342792).

Saliva DNA collection

Saliva collection was particularly suited for postal recruitment of subjects. Each saliva collection kit was clearly labelled with the subject's name and unique ID. Two makes of saliva kits were used, the Oragene OG-250/OG-500 or the Isohelix GFX-02 4ml GeneFix. In subjects who had difficulty spitting, the Oragene OG-575 DNA swab collection kit was used (see Figure 6.1).

Figure 6.1. Oragene OG-500 (left) and OG-575 (right) saliva collection kit.

DNA extraction protocols

Purification of DNA from whole blood was performed using the QIAamp Blood Maxi Kit (Spin Protocol). In whole blood, DNA is extracted from white blood cells. Saliva DNA extraction for the Oragene OG 250, OG 500 and OG 575 was performed using the PrepIT L2P kit. Saliva DNA extraction protocol for Isohelix GFX-02 4ml GeneFix was performed using the GSPN-12 GeneFix Full DNA Isolation Kit. The DNA collected from saliva originates from buccal epithelial cells and from white blood cells. Appendix F details the specifics of the DNA extraction protocols used.

Measurement of DNA

Nucleic acids absorb ultraviolet light due to the heterocyclic rings of nucleotides. The wavelengths of maximum absorption are 260nm and 280nm respectively. The ratios of absorbance at these wavelengths A260/A280 are used as a measure of purity with a ratio of 1.8 generally accepted as "pure" for DNA. The Nanodrop ND 100 Spectrophotometer was used to check the A260/A280 wavelength ratio and DNA concentration of samples.

6.2 Exome sequencing

The human genome consists of three billion nucleotides which make up DNA sequences. These DNA sequences can be classified into regions known as the exon or intron. The exon derived from "expressed region" are sequences that are translated or expressed as proteins, and the intron from the "intragenic region" are sequences that are not represented in the protein. The exome comprises all of the genome's exons and makes up about 1.5% of the human genome.

The exome is a rich source of rare disease related variants and holds 85% of disease causing mutations in mendelian disorders[142,143]. Exome sequencing is a powerful technique used to selectively capture and sequence the exons in the genome. It is a particularly practical strategy as it is fast and cost effective and interrogates only the coding regions of a genome, reducing the volume of genomic data yet maintaining practicality in gene discovery.

6.3 Generation of the exome sequence

The main steps of the sequencing process can be divided into (1) library preparation and exome enrichment, (2) cluster generation and amplification and (3) sequencing and base calling. The Guys Genomic Facility performed these steps. DNA libraries were prepared from 3 µg dsDNA using the SureSelect Human All Exon 50 Mb Kit (Agilent). Samples were multiplexed

(four samples on each lane) and 100bp paired end sequencing was performed on the Illumina HiSeq system. I shall summarise the main principles of these steps in the next section.

(1) Library preparation and exome enrichment. 3µg in 150µl (20ng/µl) of DNA was required. The DNA is fragmented by ultrasonic shearing to 150bp fragments. The sheared blunt ends are repaired and adenylated. Adaptor oligos are ligated to both ends of the fragment and the DNA fragments are size selected and purified. Exome capture occurs by hybridization with biotinlyated RNA oligos for 24 hours. An exome library is created after removal of untargeted regions.

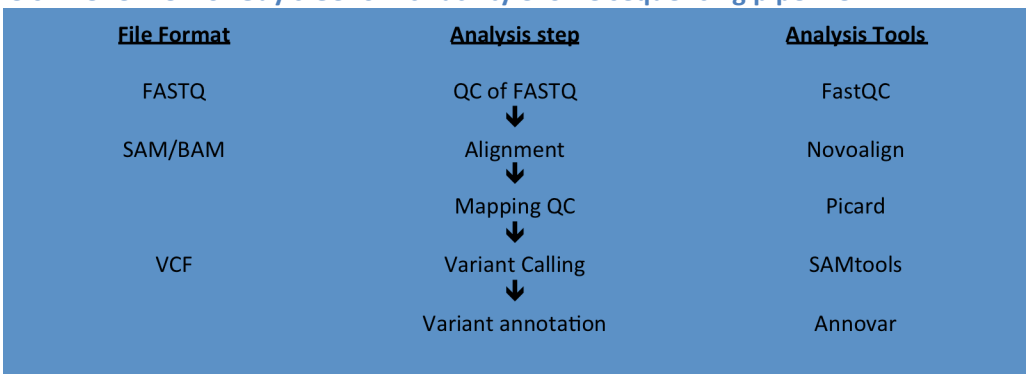
(2) Cluster generation and amplification. Sequencing occurs in arrays called flow cells. Each flow cell contains eight independent lanes during a single sequencing run and each lane contains four samples that are indexed. Each flow cell has sequencing primers that bind to the adaptor oligos at both ends of the DNA fragment. The DNA fragments are extended by bridge amplification and cloned to form clusters. This makes clonal DNA. A mismatch error of about 25% can occur which is discarded.

(3) Sequencing and basecalling. Hundreds and millions of clusters of DNA fragments are sequenced simultaneously and in parallel with reversible terminator-based method. The DNA fragment is sequenced base-by-base using four fluorescent-labelled reversible binding nucleotides (dNTPs). The four nucleotides compete with each other to bind to the DNA template. Since all four reversible terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias. After each round of synthesis, a laser is used to image and emit the colour of the newly added base. The dNTP is then cleaved to allow incorporation for a newly added base. The steps are then repeated to extend the strand and generate base calls. Base calling algorithms infer the actual nucleotide information from the obtained fluorescence-intensity data for each cluster of DNA templates. A measure of uncertainty (quality score) is assigned to each base call.

6.4 Exome sequencing pipeline

Data analysis in exome sequencing involves several bioinformatics analysis steps. Different commercially available and open access tools are available for each stage. Herein the tools presented are those used by the Guys's Genomics Facility in their exome sequencing data analysis pipeline, and which was used to generated the data described here. Figure 6.2 gives an overview of this pipeline and the process is described in detail below.

Figure 6.2. Overview of Guy's Genomic Facility exome sequencing pipeline.



QC quality control, VCF variant call format, SAM sequence alignment file, BAM binary alignment map

6.4.1 Quality Control of raw sequencing data

In general, quality control (QC) steps are carried out to assess quantity of data, base quality score distribution and base content in order to check the standard of the sequencing data. During base calling, a phred quality or Q score is assigned for each base as a prediction of an incorrect base call. Given a base call X, the probability that X is not true P(~X) is expressed by $Q(X)=-10\log_{10}P(\sim X)$. Therefore, a quality score of 10 indicates an error probability of 0.1 and a base call accuracy of 90%. Quality control of raw sequencing was carried out by FastQC (http://hannonlab.cshl.edu/fastx_toolkit).

The QC steps are as follows: (i) Illumina purity and chastity filters - cloned reads can become out of sync during sequence incorporation of bases, a mixed signal is detected and the reads excluded; (ii) homopolymer filter to check if the same base is called all the way through; (iii) polyclonal filter to remove reads if 8 of the first 20 bases have Q score <10, (iv) minimum information filter to ensure each read has >20 bases with a Q score >30, (v) remove adaptor sequences from 3' end and (vi) base composition assessment particularly %GC content which may be biased due to polymerase chain reaction (PCR) steps during library preparation.

Using CASAVA, an Illumina platform software, the raw sequencing files are converted into a FASTQ format. FASTQ is a text-based format that is used to store unique sequence ID, nucleotide sequence and its corresponding quality score. At this stage there are about 80,000 raw unaligned reads.

6.4.2 Alignment

Alignment is the process of mapping short nucleotide reads to a reference human genome and is considered the most important step in the exome sequencing pipeline. This was carried with Novoalign (http://www.novocraft.com). Novoalign deals with gaps in alignment, strips off the adaptor sequences, and performs quality score and quality re-calibration.

The resultant sequence alignment is stored as a Sequence Alignment Map (SAM) file. This is an uncompressed tab delimited text file for storing generic large nucleotide sequence alignment format where essential alignment information such as mapping position is recorded. A Binary Alignment Map (BAM) format is the binary representation of SAM and keeps exactly the same information. It is a compressed and indexed file. A BAM file is technology independent and can contain data from a single or several samples.

6.4.3 Post alignment and variant calling

Sequencing error and PCR can result in propagation of duplicates. Picard tools, a Java-based command-line was used to manipulate SAM files and marks and removes duplicates.

The next step is variant calling, which is carried out by comparing the aligned sample sequences with known reference sequences to determine which positions deviate from the reference. Variant calling can be complicated by alignment and sequencing errors such as the presence of indels, PCR artefacts, variable GC content in short reads and variable quality scores. SAMtools was the program used for variant calling. It has a Bayesian variant and short indel caller and can sort, remove cluster single nucleotide polymorphisms (group of SNPs mapped to an identical location), merge, index and generate alignment and create consensus sequences. Variant calling QC involves calculation of variant metrics, transition/tranversion ratios, indel size distribution, capture efficiency and completeness of coverage. This produces a list of variant calls recorded as a variant call format (VCF).

VCF is a generic format for storing DNA polymorphism data, together with rich annotation. The format consists of a series of meta information lines, header lines and then data lines each containing information about a position in the genome.

6.4.4 Gene and variant annotation

Annotation is the process where variants in a VCF file are tagged or annotated with information from various databases in order to inform specific characteristics of the variant such as minor allele frequency (MAF) or functional consequence to the associated protein. This helps with informed variant filtering during analysis. Annovar was used to perform gene based and functional annotation of genetic variants for human genome build GRCh37 (hg19) http://annovar.openbioinformatics.org/en/latest/user-guide/startup/[144]. Annovar uses an input format with details of chromosome, start position, end position, reference nucleotide and observed nucleotide. The relevant functions and outputs obtained were:

- Gene based annotation: identify whether the SNP cause protein coding changes and amino acids that are affected.
- Filter based annotation: identify variants that are documented in the 1000 Genomes Project, Exome Variant Server or Exome Aggregation Consortium. Calculate SIFT, PolyPhen, CADD and splice site predictors.

The exome sequencing annotated output file was presented in a tab delimited text file with 28 columns. Table 6.1 details the description of the individual outputs derived from the Guy's Genomics Facility pipeline. In addition to this, I used Ensembl Variant Effect Predictor (www.ensembl.org/vep) to annotate and determine effect of variants of interest.

Table 6.1 Description of annotated exome sequencing output file.

Column	Example output	Description
1	S0633	Sample ID
2	chr10	Chromosome 10
3	101421279	Start position
4	101421279	End position
5	G	Reference allele
6	A	Alternate allele
7	HET	Heterozygous or homozygous call
8	exonic	Exonic or intronic call
9	ENTPD7	Gene name
10	Nonsynonymous	Functional annotation
11	SNV	Single nucleotide variant
12	ENTPD7:NM_020354:exon3:c.G85A:p.V29M	gene:NM_ID:exon_no:cDNA_change:protein_change
13	.	
14	RS35229854	Reference SNP cluster ID (rsID)
15	0.0012	NHLBI ESP project allele frequency
16	0.0034	ExAC frequency
17	0.00199681	1000 Genomes project frequency
18	18.67	CADD score
19	0.24	SIFT score
20	0.911	PolyPhen2hdiv score
21	0.491	PolyPhen2hvar score
22	15_hom	Number of homozygous in-house samples with this variant
23	9_het	Number of heterozygous in-house samples with this variant
24	PASS	Passed the quality filter
25	171	Quality score or Q score is a prediction of an incorrect base call. Given a base call X, the probability that X is not true P(~X) is expressed by Q(X) = -10log$_{10}$((P(~X)) where P(~X) is the estimated probability of the base call being wrong. A quality score of 10 indicates an error probability of 0.1.
26	DP=85;VDB=0.0399;AF1=0.5;AC1=1;DP4=37,6,35,5;MQ=60;FQ=174;PV4=1,0.061,1,0.037	(DP) Read depth; (VDB) Variant Distance Bias; (AF1) Max-likelihood estimate of the first ALT allele frequency; (AC1) Max-likelihood estimate of the first ALT allele count; (DP4) high-quality ref-forward bases, ref-reverse, alt-forward and alt-reverse bases; (MQ) Root-mean-square mapping quality of covering reads; (FQ) Phred probability of all samples being the same; (PV4) *P*-values for strand bias, baseQ bias, mapQ bias and tail distance bias.
27	GT:PL:DP:SP:GQ	GT (genotype, e.g. 0/1 is ref/first alt): PL (List of phred-scaled genotype likelihoods): DP (read depth): SP (Phred-scaled strand bias *P*-value): GP (Genotype Quality)
28	0/1:201,0,213:83:0:99	Values for each abbreviation listed in column 27, separated by ":".

6.5 CNV detection

A copy number variation (CNV) is a structural variation caused by a duplication or deletion leading to a variation in the number of copies of DNA. CNVs are called from exome sequencing data using the expected proportion of depth of coverage at a particular position. The average read counts are calculated at any particular exome region and CNVs are inferred from variation

in this read depth. A R package software called ExomeDepth was used[145]. ExomeDepth was programmed to feed 100 controls both ways from the sample ID and searches for the control samples with the best coverage distribution match of read count data with the sample. The program is also able to mark whether the CNV is in a known CNV area or in a region of segmental duplication.

6.6 Limitations of exome sequencing

First, and possibly the most important is its fundamental limitation in the absence of coverage in potentially functional non-coding elements. This technique presumes that disease causing mutations only occur in protein coding regions by excluding interrogation of intronic mutations, promoter regions, untranslated regions, enhancers and noncoding RNAs. These regions can harbour pathogenic variants although this contribution in monogenic disorders is likely to be small[146].

Next, exome sequencing relies on defined protein coding regions in the genome but the definitions of these regions are constantly evolving. The first exome capture kits were designed based on the consensus coding sequence consortium database (CCDS). The Agilent CCDS SureSelect Human All Exon Kit covers 37.6Mb and 30030 genes as compared with the GENCODE exome released in 2011 with a coverage of 47.9Mb and 35989 genes[147]; moreover an updated GENCODE exome is currently being prepared. Therefore, certain protein coding regions might be incompletely annotated and coverage of coding regions is platform dependent.

The efficacy of exome sequencing is dependent on the sequence capture probes and sequencing platform used. Uneven uniformity of coverage and uneven read depth from differential capture efficiencies where reads are either on, off or near target annotated regions can lead to false positive and false negative calls in genotyping. False calls are also commonly a result of misalignments, particularly as segmental duplications; intervals that are larger than 1kb with homology of >90% with other parts of the genome, encompass 5% of the human genome, can lead to misalignment of reads to the reference genome due to these similar DNA sequences. Batch effects in different sequencing runs further complicate this by creating subtle different patterns of signal depending on which genomes are analysed together. Therefore, Sanger sequencing via PCR reactions remains the gold standard for verification of variants identified by exome sequencing.

Lastly, exome sequencing has a limited ability to detect structural variations due to the non-contiguous nature of the captured exons. It may not capture medium or large indels, repeats, copy number variants, translocation and inversions. Several software packages have been developed to address this, although the robustness of these techniques is still being debated[148].

6.7 Annotation methods

Prediction tools

Variants are annotated for their potential pathogenicity using multiple algorithms and systems. These predictors have individual strengths and weaknesses and when six predictors (CADD, SIFT, Polyphen-2, Provean, Mutation Taster and UMD-Predictor) were tested with 5000 variants from Uniprot, they reached consensus for only 3074 (61.5%) of variants, while when using only the two most efficient systems (CADD and UMD Predictor), consensus was achieved for 4275 (85.5%) variants[149]. Thus using multiple prediction tools can be counterproductive and three tools, SIFT, Polyphen-2 and CADD were selected. SIFT and Polyphen were selected, as they are historical, frequently used tools, which are assessable to readers. CADD is a consistently good performing predictor and have been validated in its use in epileptic encephalopathy[150]. In addition to specific variant prediction, gene filtering was aided with the use of RVIS scores and ExAC Z scores, which will be discussed later.

6.7.1 Sorts Intolerant From Tolerant

SIFT is a sequence homology based tool that **s**orts **i**ntolerant **f**rom **t**olerant amino acid substitutions. It predicts whether an amino acid substitution affects protein function and can be applied to naturally occurring nonsynonymous polymorphisms or laboratory-induced missense mutations[151]. SIFT is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences collected through PSI_BLAST (Position-Specific Iterated Basic Local Alignment Search Tool). This relies on the presumption that amino acid residues that are essential for protein function should be evolutionally conserved. SIFT is available as an online tool (http://sift.jcvi.org). SIFT provides a prediction cut-off score of 0.05. Scores ≤0.05 are deemed damaging and >0.05 are deemed tolerated.

6.7.2 Polymorphism Phenotyping v2

PolyPhen-2 (Polymorphism Phenotyping v2) predicts the possible impact of an amino acid substitution on the structure and function of a human protein[152]. The tool uses features comprising the sequence, phylogenetic and structural information characterising the substitution. An amino acid replacement might occur at a site where binding to other molecules or formation of a secondary or tertiary structure is disrupted e.g. disulphide bond, transmembrane domain etc. The HumVar-trained PolyPhen-2 (PP2hvar) model was used for this study. The HumVar datasets consist of all human disease causing mutations from UniProtKB, together with common human non-synonymous SNPs (MAF>1%) without annotated disease involvement, which was treated as non-damaging (http://genetics.bwh.harvard.edu/pp2/dokuwiki/docs). For a mutation, PolyPhen-2 calculates a naïve Bayes posterior probability that the mutation is damaging. An appraisal of benign (score 0 to 0.446), possibly damaging (score 0.447 to 0.908) and probably damaging (score 0.909 to 1) of each variant is given.

6.7.3 Combined Annotations Dependent Depletion

Combined Annotations Dependent Depletion (CADD) scores for deleteriousness of SNPs and small insertions and deletions (indel) variants in the human genome[153]. CADD integrates 63 annotations into one metric and calculates deleteriousness by contrasting 14.7 million variants that survived natural selection with 14.7 million simulated mutations. Annotations span a range of data types, including conservation metrics, regulatory information, transcript information, cell lines expression levels and protein level scores such as those generated by SIFT and PolyPhen. A C score is obtained and correlates with allelic diversity, annotations of functionality, pathogenicity, disease severity, experimentally measured regulatory effects and complex trait associations. Pre-computed CADD based scores (raw C-scores) are available for all 8.6 billion possible SNPs in the reference genome, 1000 Genome and Exome Sequencing Project (ESP) variant releases. Raw C scores can be phred-scaled by expressing the rank in order of magnitude to scaled C scores. For example, a scaled CADD score of 10 indicates the top 1% of deleterious variants and CADD scaled 20 the top 0.1% deleterious variants. These scaled C scores are used to easily filter out the top proportion of damaging variants. A CADD score of 10 was used as the cut off for the annotation analysis of known variants and a score of 20 during gene exploration with aetiologically relevant gene sets.

6.7.4 The Residual Variation Intolerance Score

The Residual Variation Intolerance Score (RVIS) denotes how intolerant a particular gene is to genetic variation by comparing observed and predicted levels of standing variation in the regulatory region of human genes (http://genic-intolerance.org). Genes are given a percentile RVIS score, genes that are more intolerant and hence less likely to demonstrate genetic variation will have a lower RVIS percentile and are more suggestive to be pathogenic. RVIS appears to be a sensitive measure for use in EE and when applied to the trio sequencing dataset of Epi4K cohort, *de novo* mutations were observed to lie in the most intolerant 25th percentile RVIS scores when compared with controls or immunological disorders[154]. In addition, a RVIS ≤25th percentile was used to assess the likelihood of pathogenicity of variants in a further study of 356 EE trios[155] and this tool was used to prioritise gene variants for my analysis (see section 6.15.1)

6.7.5 Splice site annotation tools

The Annovar system provides splice site effect predictions by AdaBoost (Ada) and random forests (RF) (section 6.5). Ada and RF are ensemble learning methods that were constructed following the evaluation of eight in silico splice site tools[156]. The eight tools examined were Position Weight Matrix model, MaxEntScan, Splice Site Prediction by Neural Network, GeneSplicer, Human Splicing Finder, NetGene2, GENSCAN and SplicePredictor. The authors restricted the testing dataset to SNVs within splicing consensus regions (-3 to +8 at the 5' splice site and -12 to +2 at the 3' splice site), and defined a variant to be splice altering if either its Ada score or RF score was larger than 0.6. Splice site annotation was not available through the Guy's Genomics Facility pipeline and splice variants of interest were annotated separately using Annovar.

6.7.6 Gene expression

Gene expression was determined using genevestigator (https://genevestigator.com). Genevestigator is a high performance search engine for gene expression that integrates thousands of manually curated, well-described public microarray and RNAseq experiments and visualises gene expression across different biological contexts. Gene expression in the gene of interest (GOI) was reviewed in human central nervous system tissues.

6.7.7 Gene and variant association with disease

Several different resources were used to determine gene and genetic variant association with disease. These resources included Pubmed search, the Online Mendelian Inheritance in Man (OMIM), Genecard, Clinvar, Uniprot and HGMD. In some instances locus specific databases such as http://www.molgen.ua.ac.be/scn1amutations/Home/Default.cfm for *SCN1A* and http://mecp2.chw.edu.au for *MECP2* and *CDKL5* and mutations were used.

6.8 Integrative Genomics Viewer

The Integrative Genomics Viewer (IGV) is used to visually explore sequencing reads at various resolutions from whole chromosome to base level[157]. A password protected BAM file was loaded from http://athena.kcl.ac.uk/igv-data/exome/SAMPLEid/SAMPLEid_novoalign.bam and the region of interest visually interrogated. Visual interrogation was used to view coverage data, read depth, individual aligned reads and individual base mismatches which are highlighted by colour and intensity according to base call and quality. IGV interrogation of variants of interest was used to filter out probable false positives from sequencing data (described in section 6.17). IGV plots of variants of interest analysed are listed in appendix I.

6.9 Control population sequencing databases

Four open source publicly available control population databases were used. These were the 1000 Genomes Project (1000G), the Exome Variant Server (ESP), the Exome Aggregation Consortium (ExAC) and the Genome Aggregation Database (gnomAD).

The 1000 Genome Project

The 1000 Genomes Project is a public catalogue of human genetic variation including SNPs and structural variation. The goal of the 1000 Genomes Project was to find most genetic variants with frequencies of at least 1% in the populations studied[158]. The resource consists of 2504 individuals from 26 populations. All participants declared themselves to be healthy at the time that the samples were collected. Each sample was sequenced to 4x genome coverage. Low coverage and exome sequence data are available.

The Exome Variant Server

The goal of the Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP) was to discover novel gene and mechanisms contributing to heart, lung and blood disorders[159]. The current release (ESP6500SI-V2) consists of 6503 individuals (2203 unrelated African Americans

and 4300 unrelated European Americans) recruited from North American Institutes with heart, lung and blood disorders. Specifically neurological and psychiatric disorders were not considered. The ESP exome variant data were aligned to GRCh37 using Genome Analysis Toolkit (a tool for processing and variant calling high throughput sequencing data). SNP calls were generated using the Michigan SNP calling pipeline (http://genome.sph.umich.edu/wiki/UMAKE). Small indel variants were analysed at the Broad Institute using the GATK variation discovery pipeline following the guideline in the GATK best practices v4. SeattleSeqAnnotation137 annotated all SNPs and indels (http://snp.gs.washington.edu/SeattleSeqAnnotation137).

The Exome Aggregation Consortium

The Exome Aggregation Consortium (ExAC) seeks to aggregate and harmonise exome sequencing data from a variety of large-scale sequencing projects, including some data from the 1000 Genomes and ESP, in order to make the data available to the wider scientific community[160]. The resource consists of 60706 unrelated individuals as part of various disease specific and population genetic studies. Individuals affected by severe paediatric disease are removed. All the raw data from different projects were reprocessed through the same pipeline and jointly called to increase specificity.

In addition to informing MAF, the ExAC Z score for missense mutations was used when prioritizing candidate genes. A Z score represents a score of tolerance and indicates a deviation of observed from expected counts.

The Genome Aggregation Database

The Genome Aggregation Database (gnomAD) is the latest release from the ExAC group (http://gnomad.broadinstitute.org). Rather than containing only exome data with ExAC, gnomAD dataset spans 126,216 exomes and 15,137 genomes. Individuals that were known to be affected by severe paediatric disease as well as their first-degree relatives have been removed from the dataset. Raw data from the different sequencing projects have been reprocessed through the same pipeline and jointly variant called to increase consistency across projects. However, multiple exome capture methods and sequencing chemistiries were used so there was variable coverage between individuals and across project sites.

gnomAD is currently in early beta mode (November 2016) and was not part of the standard exome sequencing analysis pipeline. However, all novel and candidate gene variants shortlisted were manually re-annotated with a gnomAD MAF.

6.10 Gene identification strategies

The era of exome sequencing has brought about abundant genetic data. Beyond the stringent technicalities from sequencing production to variant annotation, the challenge in gene discovery lies in the filtering of genetic variants based on principles underlying the genetic model of the disease and the cohort available. Various disease gene identification strategies in exome sequencing have been described[98], and the use of association testing can add statistical power[155,161].

A shared rare variant approach appears intuitive, but in MAE this situation is compounded by limited power due to genetic heterogeneity and sample size restrictions due to its rarity. The current genetic understanding of MAE indicates that approximately 20% of cases are linked to a monogenic aetiology spanning 10 genes[9-23]. In the majority of cases, reported pathogenic variants are *de novo*, although private mutations in multiplex families have been reported as well as parent to child affectedness[17]. In addition, the emerging paradigm of the role of *de novo* mutations' impact in developmental disorders and neuropsychiatric phenotypes was acknowledged[76,162]. The discovery strategy would lean towards trio sequencing for identification of *de novo* variants but due to a limited financial budget, the exome sequencing of trios would restrict the actual number of cases being sequenced and in turn reduce the yield of discovery. The approach with a potentially more promising yield and that I decided to use in this project, was to exome sequence cases, and then prioritise genetic variants based on different filtering criteria and associated hypothesis. Candidate variants assembled were validated using PCR and Sanger sequencing, with inheritance studies performed where possible. The different bioinformatics filtering approaches can be divided into the following:

(1) Annotation analysis of reported epilepsy associated genetic variants.
(2) Case control association analysis.
(3) *De novo* analysis of Euroepinomics exome sequencing trios.
(4) Review of known epilepsy associated genes novel variants.
(5) Gene exploration with aetiologically relevant gene sets.
(6) Shared novel variant analysis of unsolved cases.
(7) Analysis of sibling pair families.

6.11 Annotation analysis of reported epilepsy associated genetic variants

Annotation analysis of reported epilepsy associated gene variants was performed in order to provide guidance of the thresholds and limits to set for MAF and with *in silico* predictors

during analysis and filtering of exome sequencing results. Epilepsy variants were assembled from the three databases below (accessed September 2015).

(1) ClinVar https://www.ncbi.nlm.nih.gov/clinvar/

A search term of 'epilepsy', 'epileptic encephalopathy', 'myoclonic epilepsy' and 'myoclonic astatic epilepsy' was used to assemble relevant variants. Variants annotated as benign and translocations were removed. Two datasets were then created, first the 'All_epilepsy' group that consisted of heterogeneous groups of disorders with epilepsy symptoms included in a diagnosis such as ID syndromes, metabolic disorders and congenital syndromes etc. The second 'Pure_epilepsy' group where each Clinvar entry was reviewed and only classified epilepsy syndrome[29] variants were kept. Additionally in the Pure_epilepsy group, the epilepsy syndromes and genes were stratified into autosomal dominant or autosomal recessive groups by individual literature review of published variants. These ClinVar variants were converted from hg20 to hg19 using UCSC lift over tool https://genome.ucsc.edu/cgi-bin/hgLiftOver.

(2) The Human Gene Mutation Database. https://www.hgmd.cf.ac.uk

This is a curation of published gene lesions responsible for human inherited disease, with the public release containing 121,002 total mutations. A search term of 'epilepsy' was used. All variants identified were included in the All_epilepsy group described above.

(3) EpilepsyGene. http://61.152.91.49/EpilepsyGene.

This is an open access online database containing 499 genes and 3931 variants associated with 331 clinical epilepsy phenotypes collected from 818 publications curated by researchers[163]. All variants in the database were included into the All_epilepsy group as the investigators selected epilepsy variants from heterogenous epilepsy related studies including metabolic and mitochondrial disease.

All entries were collated together within the two groupings, duplicates were removed and entries were converted into Browser Extensible Data (BED) file format (column headings: chr start-end ref alt). Annotation was performed using Annovar (see appendix E for annotation script). The main annotated output fields of interest were: MAF based on 1000 Genomes, ESP, ExAC and *in silico* predictions with SIFT, PolyPhen-2hvar and CADD. In analysing the range of scores, pathogenicity cut off scores for SIFT and PP2 was used as previously indicated and CADD scores above 10 was considered.

MAF for each variant was determined based on which population databases the variant was recorded in, and then weighted according to the sample size of the population dataset. When the variant was not seen in any database, it was considered novel; when present in one database this value was used, when present in two or three databases a weighted MAF based on the samples size in the respective database was used. Table 6.2 shows the weighted factors used to calculate the final MAF.

Table 6.2 Minor Allele Frequency weighted factors in the three population databases.

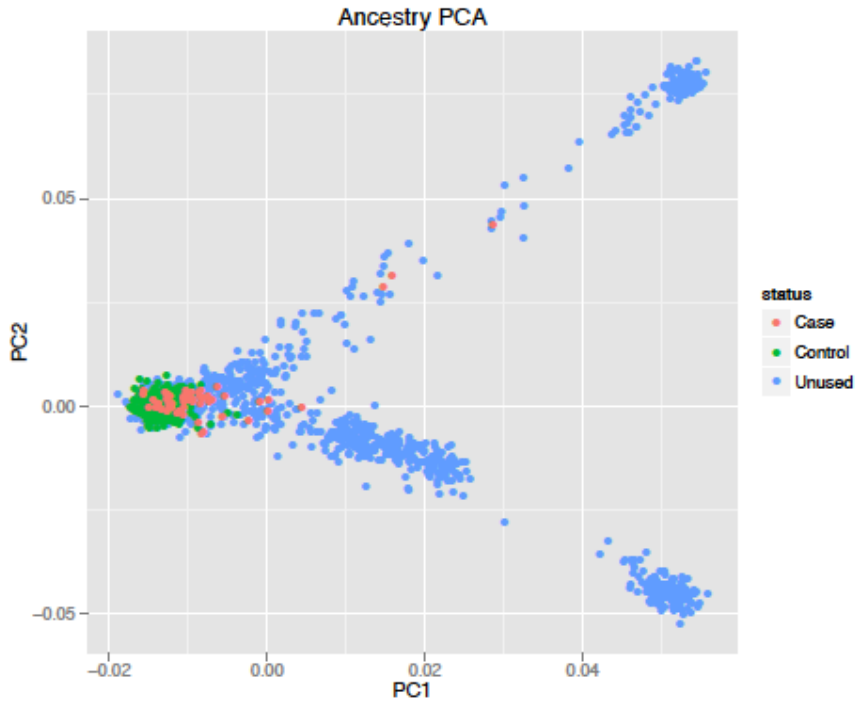
1000 Genomes (n=2403)	3.59	27.8	3.96	
ESP (n=6503)	9.33	72.2		9.68
ExAC (n=60706)	87.08		96.04	90.32

6.12 Case control association analysis

One of the initial methods used to identify causative variants was a gene based rare variant case control association analysis. This was performed using EPACTS v3.2.3 (Efficient and Parallelizable Association Container Toolbox) (http://genome.sph.umich.edu/wiki/EPACTS) and plotting was performed with R 3.0.0. EPACTS is a framework that allows for gene based or single variant statistical association analysis using NGS data. Professor Michael Simpson developed the script that was used.

Controls were randomly selected from in house exome sequencing samples. Cases or relatives of cases with epilepsy were previously removed. First, a principal component analysis (PCA) plot was generated to determine whether cases and controls were ancestrally matched. PCA reduces SNPs of common variants not in linkage disequilibrium with each other and sample data to linear uncorrelated representation called principal component (PC). Each PC is orthogonal to other PCs and is ranked by variance. The first PC has the largest possible variance and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. 55 cases and 524 controls were used in this comparison. The script calculated the five closest controls. Figure 6.3 shows the ancestry PCA plot of cases and controls.

Figure 6.3 Ancestry PCA plot of cases and controls.



The first two principal components PC1 and PC2 are shown. Three samples (from left to right PC1 0.0147; S1381 – Italian case, PC1 0.0158; S1389 - Italian case and PC1 0.0286; S1273 – UK case, mixed Pakistani, Black Afro-Caribbean and White English) appeared as outliers and were removed from further analysis.

Three samples (S1381, S1389, S1272) were identified as outliers from the PCA plot and were removed. The remaining samples were tested with controls using four models of inheritance and annotated function; dominant altering (dom.alter), dominant truncating (dom.trunc), recessive altering (rec.alter) and recessive truncating (rec.trunc). Altering variants encompassed nonsynonymous missense variants whereas truncating variants were nonsynonymous frameshift, stop gain and splice site variants. A gene collapsing fisher test was used, where the number of rare variants (set at MAF <0.01 for dominant and MAF <0.05 for recessive) in any one gene was counted in cases and compared with controls. A quartile-quartile plot demonstrating the distributions of the expected versus observed *P* values for any given rare variant was also plotted. A quartile quartile plot is a plot of the quantiles of two distributions against each other, the pattern of points in the plot is used to compare the two distributions and a visual goodness of fit assessment. The ten genes obtained with the lowest *P* values for each model were reviewed to access suitability as a candidate gene.

6.13 *De novo* analysis of Euroepinomics MAE trios

The Euroepinomics consortium assembled 38 MAE trios for exome sequencing. Sequencing was performed at the Wellcome Trust Sanger Institute (Hinxton, Cambridgeshire) and has

been described before[77]. In brief, paired end DNA librarys were prepared using TruSeq DNA Sample Preparation Kit from Illumina, target enrichment performed with SureSelect Human All Exon 50Mb Kit from Agilent Technologies. DNA was sequenced on a HiSeq2000 Illumina as paired end 75bp reads. Sequencing reads were aligned using Burrows Wheeler Aligner (http://bio-bwa.sourceforge.net/). GATK was used to perform QC on Fastq files. Variant calling was performed with SAMtools mpileup, GATK, UnifiedGenotyper (http://software.broadinstitute.org/gatk/) and Dindel[164]. GenomeComb (http://www.genomecomb.sourceforge.net/) was used to annotate and filter data.

The Euroepinomics consortium carried out *de novo* variant analysis. DeNovoGear version 0.2 (http://www.denovogear.weebly.com) program by Conrad and colleagues was used to detect *de novo* SNVs and Indels from trio exome BAM files. This was double checked by GenomeComb analysis.

Sequencing files were transferred and included into the MAE exome sequencing cohort. VCF files were converted back to fastq files and variant calling re-performed using the Guy's Genomics facility pipeline.

6.14 Review of known epilepsy genetic variants

During the course of the project, the number of genes associated with MAE increased from two to ten genes uncovering the extent of the genetic heterogeneity of MAE. Whilst some genes had specific phenotypic characteristics (e.g. *SCN1A*); many other genes showed diverse phenotypic characteristics (e.g. *TBC1D24*). Hence, the list of genes that was interrogated was expanded from genes only associated with MAE to most epilepsy-associated genes; in order to identify whether MAE was within their phenotypic spectrum.

Variants in 100 known epilepsy genes were reviewed. The gene list assembled (April 2016) was consistent with the King's Health Partners epilepsy gene panel, and most commercially available epilepsy gene panels (Amplexa, Ambry). The list contains genes that are most commonly associated with epilepsy and includes examples of metabolic disorders (*PNPO*), neonatal encephalopathy (*KCNQ2*), specific epilepsy syndromes (*GRIN2A*), ID syndromes (*IQSEC2*), congenital syndromes (*MECP2*), cortical malformations (*MTOR*) as well as specific MAE associated genes (*SLC6A1*). The epilepsy associated genes reviewed are listed in Table 6.3.

Table 6.3 List of known epilepsy genes.

ADSL	*DNM1*	*HNRNPU*	*PIGA*	*SLC1A2*
ALDH7A1	*EEF1A2*	*HUWE1*	*PIGO*	*SLC25AA*
ALG13	*FASN*	*IQSEC2*	*PIGT*	*SLC2A1*
ARHGEF9	*FOXG1*	*KCNA2*	*PIK3AP1*	*SLC35A2*
ARX	*GABBR1*	*KCNB1*	*PIK3R2*	*SLC35A3*
ATP1A2	*GABBR2*	*KCND2*	*PLCB1*	*SLC6A1*
ATP1A3	*GABRA1*	*KCNH5*	*PNKP*	*SLC6A8*
ATRX	*GABRA5*	*KCNQ2*	*PNPO*	*SLC9A6*
CACNA1A	*GABRB3*	*KCNQ3*	*POLG*	*SMARCA2*
CACNA1D	*GABRD*	*KCNT1*	*PRRT2*	*SPTAN1*
CACNA1H	*GABRG2*	*KIAA2022*	*PURA*	*ST3GAL3*
CACNB4	*GAMT*	*LAMC3*	*RELN*	*STX1B*
CDKL5	*GATM*	*LGI1*	*SCN1A*	*STXBP1*
CHD2	*GNAO1*	*MECP2*	*SCN1B*	*SYNGAP1*
CHRNA2	*GRIN1*	*MEF2C*	*SCN2A*	*SZT2*
CHRNA4	*GRIN2A*	*MTOR*	*SCN8A*	*TBC1D24*
CHRNB2	*GRIN2B*	*NBEA*	*SETBP1*	*TCF4*
CNKSR2	*GRIN2D*	*NPRL2*	*SIK1*	*UBE3A*
CPA6	*HCN1*	*NPRL3*	*SLC12A5*	*WWOX*
DEPDC5	*HDAC4*	*PCDH19*	*SLC13A5*	*ZDHHC9*

Only novel variants i.e. variants not found in 1000G, ESP and ExAC were considered. Deleterious variants have lower MAF than neutral ones due to negative selection, as MAE is a rare severe childhood epilepsy, MAF was expected to be rare or ultra rare (MAF <0.0001). Indeed, annotation analysis from the Pure_epilepsy variant group (see section 6.14) were mostly novel. Novelty was also a feature in the solved cases in the Euroepinomics cohort. Following filtering for novelty, variants from MAE cases that were annotated as synonymous and non-frameshift were removed. Each variant was then reviewed on IGV and classified into whether the variant was failed, was inconclusive or passed based on visual interpretation of the sequencing region surrounding the variant of interest. The criteria for IGV classification was as follows:

Failed:

- Variant call in region where total read depth less than 10 reads.
- If variant call was in less than 20% of total reads for heterozygous mutations.
- Identical indel in more than one sample and in other in-house samples.
- SNV more than 10bp from intron/exon boundary.
- Indel more than 2bp from intron/exon boundary.

Inconclusive:

- Variant call in more than 20% but less than 30% of reads.
- Total read depth more than 10 but less than 20 reads.

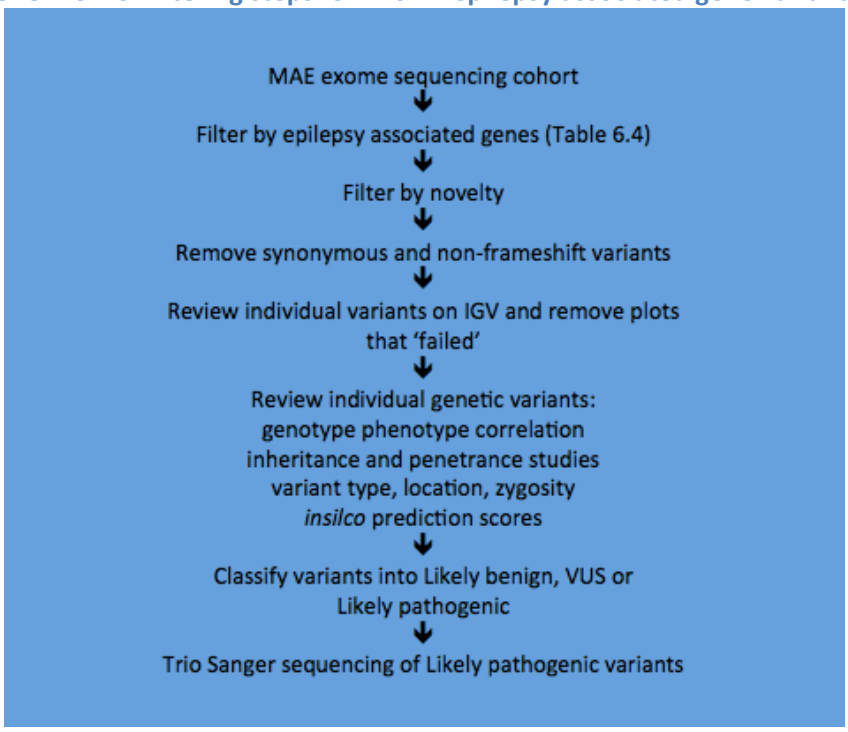
Passed:

- Everything else.

Variants that were classified failed after IGV interrogation were removed from further analysis. The remaining individual variants were than reviewed in detail.

Filtered gene variants were reviewed for specific genotype-phenotype correlation, for example if all previously reported patients with a pathogenic variants in a specific gene had severe ID and the case in question had no evidence of ID, the variant was down-weighted. Inheritance and penetrance studies were reviewed where available, and variants down-weighted where inheritance was demonstrated without evidence of phenotypic segregation. Variants where zygosity did not match mode of inheritance for the gene in question were also down weighted. The literature surrounding the genes in question were reviewed in order to determine if there was a pattern in the reported pathogenic variants e.g. location, mutation types, recurrent mutations etc. Where available locus specific databases were reviewed. Annotation with SIFT, PP2hvar, CADD and splice site (Ada and RF) *in silico* predictors were applied for each variant. Each variant was then classified into likely benign, variant of uncertain significance (VUS) or likely pathogenic. Likely pathogenic candidate variants were Sanger sequenced to validate exome sequencing results and to check inheritance in parents. An overview of the filtering steps of novel variants identified in these known epilepsy genes is shown in Figure 6.4.

Figure 6.4. Overview of filtering steps for known epilepsy associated gene variants.



De novo variants identified had maternity and paternity confirmed by selecting a rare homozygous variant in the child and Sanger sequencing the variant in the trio.

6.15 Gene exploration using aetiologically relevant gene sets

To unpick the genetic heterogeneity of MAE, gene exploration with aetiologically relevant gene sets were performed. Three gene sets, neuropsychiatric genes, ion channel genes and genes from monogenic diseases with seizures were curated as described below. The lists of genes are found in appendix H.

In this gene exploration, we screened for novel (not present in 1000 Genomes, ExAC and ESP), CADD>20 gene set variants in 79 cases of this MAE cohort. The aim was to (1) identify novel candidate genes for MAE through direct gene matching, and (2) evaluate the contribution of genes following enrichment analysis using 33,370 population controls from the ExAC Non Finnish European (NFE) dataset.

Neuropsychiatric gene set

There lies an increased awareness of shared genes, pathways and co-expression networks between the various neurodevelopmental phenotypes of autism, ID, and epilepsy[163,165-168]. The prevalence of epilepsy in patients with autism and ID is up to 40%, and conversely autism and ID are the most co-morbid conditions in epilepsy[169]. Pathogenic genes overlap and 53 candidate genes are associated with more than one disorder ($P<1 \times 10^{-6}$)[170]. Additionally, the importance of *de novo* variants in neurodevelopmental disorders is now established[101,162]. *De novo* variants in neurodevelopmental disorder tend to more prevalent[171,172], be more enriched for gene disrupting mutations (nonsense, splice site and frameshift)[172-174] and are more intolerant to functional variation with a lower RVIS percentile compared with controls[154]. Neuropsychiatric gene sets were collated from eight published *de novo* sequencing studies of four primary phenotypes, (1) developmental disorders, (2) epileptic encephalopathy, (3) intellectual disability, and (4) autism spectrum disorder (see Table 6.4). Duplicate genes were removed. The remaining 2105 unique genes with *de novo* nonsynonymous variants were curated together from all eight studies to form a neuropsychiatric gene set.

Table 6.4 *De novo* sequencing studies that contributed to the neuropsychiatric gene set.

Study	Phenotype	Subjects	*De novo* findings	No. unique genes with nonsyn variants
Developmental disorders				
DDD study[162]	Undiagnosed developmental disorders	1133 UK trios	1149 SNV and 114 Indels	1196
Epileptic encephalopathy				
Epi4K study[101]	LGS or IS	264 US trios	329 SNV	240
Intellectual disability studies				
DeLigt *et al*.[175]	IQ < 50	100 US trios	79 SNV	63
Hamdan *et al*.[171]	IQ < 50	41 Canadian trios	81 SNV	70
Ranch *et al*.[172]	IQ < 60	51 German trios	91 SNV	79
ASD studies				
Iossifov *et al*.[173]	> high functioning ASD	343 quads	754 SNVs	529
O'Roak *et al*.[176]	ASD	189 trios/quads	225 SNV and 17 Indels	181
Neale *et al*.[177]	ASD	175 trios	176 SNV and Indels	125
Sanders *et al*.[174]	ASD	200 quads	172 SNV	143

UK United Kingdom, SNV single nucleotide variation, LGS Lennox gastaut sundrome, IS infantile spasms, US United States, IQ Intelligence quotient, US United States, ASD autism spectrum disorder, nonsyn nonsynonymous

Ion channel gene set

Ion channel genes were a strongly enriched paralogue amongst *de novo* variants in 10,000 neurodevelopmental trios, with sodium channel genes conferring specific risk for severe seizures (OR 4.23) (Lal *et al.* ASHG abstract 2016). Other ion channel genes such as *KCNQ2* (potassium channels), *GABRA1* (GABA receptor channels), *CACNA1A* (calcium channels) and *CHRNA2* (cholinergic receptor channels) have all been implicated as monogenic causes of epilepsy and collectively this group accounts for a substantial proportion of epilepsy genes. 237 Ion channel genes were assembled as an ion channel gene set. They consist of all known ion channelopathy genes, their family members and functionally similar ion channels including voltage gated, ligand gated and background channels[178].

Monogenic diseases with epilepsy as phenotypic features

The third gene set selected, were genes that are known to cause monogenic diseases with epilepsy as a phenotypic feature. 335 genes were obtained from Lemke *et al.*[80].

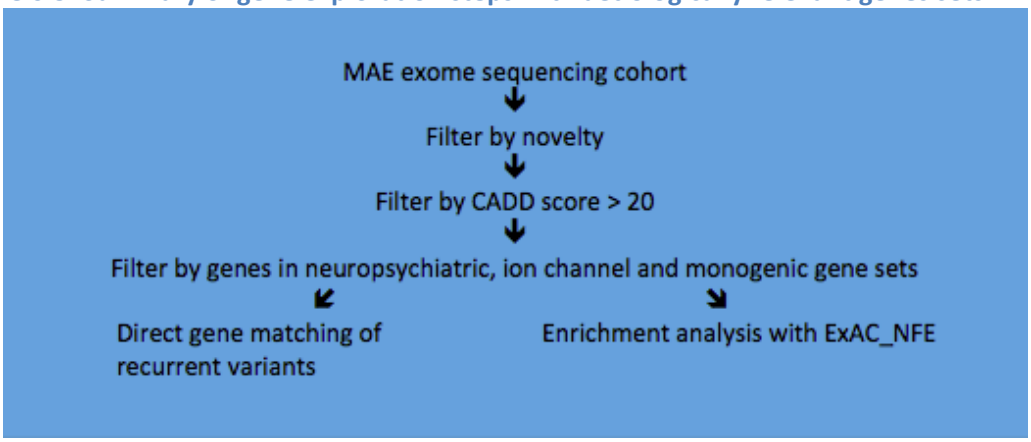
6.15.1 Direct gene matching

Genes with novel nonsynonymous genetic variant with a CADD score >20 in the MAE cohort were matched with genes form the three gene sets. In the neuropsychiatric gene set, recurrent variants where there was more than one novel nonsynonymous variant with a CADD score >20 in a matched gene were reviewed. Following each gene match, the following steps were carried out to prioritise candidate genes:

(1) Variant interrogated on IGV and classified as pass or inconclusive.
(2) GOI has a RVIS score < 25th centile or positive ExAC Z score if missense mutation.
(3) Gene expression from Genevestigator above a threshold of 'medium' in central nervous system tissue expression.
(4) Segregation data not conflicting
(5) Gene function not conflicting with MAE.
(6) Associated phenotypes not conflicting with MAE.

Genes that fulfilled all criteria were considered candidate genes and these genetic variants were Sanger sequenced in cases and parents to validate the variation and check inheritance. Figure 6.5 summaries these main steps

Figure 6.5. Summary of gene exploration steps with aetiologically relevant genes sets.



6.15.2 Enrichment analysis using chi squared analysis

Enrichment analysis per gene in each gene set was carried out using the ExAC_Non Finnish European (NFE) population cohort as controls. The ExAC_NFE cohort consist of 33,370 individuals clustered into a European population on principal component analysis with 60,706 ExAC samples.

The ExAC vcf was downloaded and converted into annovar input format. Standard annovar protocol was carried out to annotate the entire dataset. First, only novel genetic variants in genes from the three gene sets were selected in this MAE cohort. Then in the ExAC group, variants that occurred only in one individual and were not seen in 1000 Genomes project and ESP was considered novel. Next, the novel variants in both cohorts were filtered for CADD scores >20 variants. The number of filtered variants per gene was counted in each cohort and compared using chi squared test with yates correction. Level of significance was corrected for multiple testing depending on the size of the gene set. For the neuropsychiatric gene set (n=2105 genes), this was set at $P<2 \times 10^{-5}$. For the ion channel gene set (n=237 genes), this was set at $P<2 \times 10^{-4}$ and for the monogenic disease with epilepsy gene set (n=335), this was set at $P<1 \times 10^{-4}$.

The ten most statistically significant enriched genes in the MAE cohort were reviewed using the same steps as described with direct gene matching. Resultant candidate genetic variants were Sanger sequenced in cases and parents.

6.16 Shared novel variant analysis

Novel variants that were shared amongst individuals were selected out in the MAE cohort and ranked according to genes with the most number of novel variants that were shared amongst individuals to the least. Genes harbouring the most shared variants were reviewed to assess suitability as a candidate gene.

6.17 Analysis of sibling pair families

Exome sequencing data was performed for some sibling pairs. Filtering involved shared rare variant analysis and where sequencing data from other family members were available, segregation of variants was performed. For heterozygous variants, only novel variants were considered and for homozygous variants, a MAF <0.001 was considered to account for a presumed carrier frequency in the population. Synonymous, non frameshift and variants annotated with unknown function were removed. Gene variants assembled were reviewed as previously described (see section 6.15.1). PCR and Sanger validation was carried out on candidate genes.

6.18 Molecular validation methods

6.18.1 Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) is a rapid and robust method used to amplify a selected region of the genome. The procedure uses oligonucleotide primers, deoxynucleoside triphosphates (dNTPs) and Taq polymerase to synthesise DNA strands in a chosen region of interest.

The procedure involves three successive steps, which are repeated 30 to 40 times. Step 1: Denaturing, DNA is heated to a high temperature to break the hydrogen bonds holding the two complimentary DNA strands into a single stranded DNA. Step 2: Annealing, the reaction is cooled allowing the primers to bind by base pairing to a complementary sequence on a single stranded DNA. Step 3: Amplification, four dNTPs, dATP, dCTP, dGTP and dTTP and Taq polymerase from *Thermus aquaticus* initiates the synthesis of new DNA strands that are complementary to individual target DNA strands. The direction of the synthesis of new DNA strands is towards the other primer-binding site allowing the newly synthesised DNA strands to serve as templates for new DNA synthesis resulting in an exponential increase in product.

6.18.1.1 Primer design

A primer is a short synthetic oligonucleotide, which is designed to have a reverse sequence that is complement to a region of interest in a target DNA. Forward and reverse primers flank the target sequence at the 3' end of opposite strands. Primers must be carefully designed to allow successful amplification of a DNA sequence.

Primers were designed with Primer 3 software http://www.primer3.com. The DNA sequence of interest required was extracted from the UCSC Genome Browser http://genome.uscu.edu.

In order to allow maximum complementarity, annealing and avoidance of internal base pairing, primer design was carried out using the following criteria: (1) melting temperature 58-63^{o}C, within 1-2^{o}C; (2) GC content 35-70%, optimum 50%; (3) primer length 18-30bp, optimum 22bp; (4) maximum self complimentary 4bp; (5) maximum 3' self complementary 3bp; (6) maximum pair complimentary 5bp; (7) maximum polyX run of same base 3bp; (8) maximum GC in 3' end 3bp.

The selected primer was checked with University of California Santa Cruz (UCSC) Genome Browser In-Silico PCR tool https://genome.ucsc.edu/cgi-bin/hgPcr to ensure that the genomic region was tightly matched and it did not bind to any other location. Primers were purchased as unmodified DNA oligonucleotide HPSF 0.01µmol lyophilised powder from Integrated DNA Technologies (UK) or Eurofin Genomics (UK). The primers successfully used in this project are listed in appendix J.

6.18.1.2 PCR protocols

Two PCR protocols were used. A standard PCR protocol and an optimised PCR protocol when standard PCR was unsuccessful. The optimised PCR had a higher annealing temperature and reduced $MgCl_2$ from 3mmol to 2mmol, both to increase primer annealing specificity.

Standard PCR master mix protocol:

Standard PCR master mix	**10x**
Forward primer 10µmol	3µl
Reverse primer 10µmol	3µl
10x PCR buffer	15µl
5mM dNTPs	3µl
50mM $MgCl_2$	10µl
Biotaq Taq Pol	1µl
dd H_20	64µl

Standard PCR cycling conditions:

Step 1	95°C	5 minutes
Step 2 Cycle 34 times	95°C 60°C 72°C	)) 30 seconds)
Step 3	72°C	7 minutes
Step 4	10°C	∞

Optimised PCR master mix protocol:

Optimised PCR master mix	**10x**
Forward primer 10µmol	3µl
Reverse primer 10µmol	3µl
10x PCR buffer	15µl
5mM dNTPs	3µl
50mM $MgCl_2$	6.2µl
Biotaq Taq Pol	1µl
dd H_20	67.8µl

Optimised PCR cycling conditions:

Step 1	95°C	5 minutes
Step 2 Cycle 34 times	95°C 62°C 72°C	)) 30 seconds)
Step 3	72°C	7 minutes
Step 4	10°C	∞

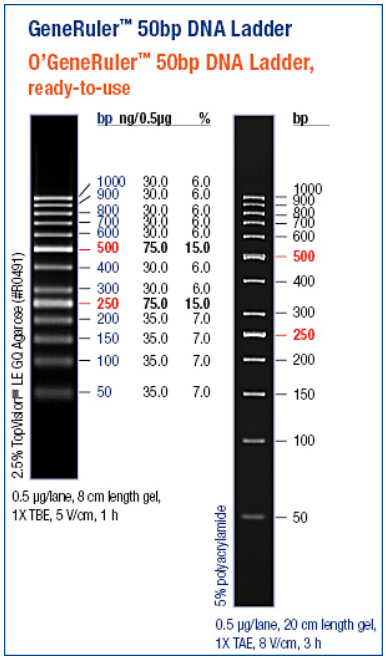
6.18.1.3 Agarose gel electrophoresis

The size of the PCR fragment can be estimated by agarose gel electrophoresis. An agarose gel acts as a sieve during electrophoresis and helps separate DNA molecules by allowing smaller DNA molecules migrate through the gel quicker than larger molecules. A 2% gel was made, as it was easier to handle and has a 100 to 2000bp range of separation. This was made by mixing 2g agarose in 100ml Tris Borage EDTA (TBE) (Sigma-Aldrich UK), and heating for about 90 seconds in a microwave on full power to dissolve the agarose. 12µl SYBR Safe DNA gel stain (Thermo Fisher Scientific (Life Technologies)) was then added to the cooled agarose TBE solution and swirled to mix. This is poured into a plate with a comb and allowed to set.

The PCR fragment is then sized with the aid of a DNA ladder and loading dye. Thermo Scientific GeneRuler 50bp DNA Ladder (Lot:0039285) was used as a DNA ladder. GeneRuler consists of a range of 13 differently sized DNA fragments from largest to smallest in order to assess the size

of the DNA fragment achieved through PCR. Figure 6.6 shows the bands of the GeneRuler 50bp DNA Ladder, it has two bright reference bands at 500bp and 250bp. It is supplied with 6x DNA loading dye. The loading dye contains two different dyes, bromophenol blue and xylene cyanol FF and helps to stain and visually track the migration of DNA fragments.

Figure 6.6 GeneRuler 50bp DNA Ladder.



7µl of solution was loaded into each gel well. For the GeneRuler, this was 1ul GeneRuler, 2µl loading dye and 4µl water. For the PCR products, this was 2µl of loading dye and 5µl of PCR product. Next, an electric charge of 100V was run through the gel from negative charge (black electrode) to positive charge (red electrode) for about 45 minutes. This is because DNA molecules are negatively charged due to the phosphates in its backbone and when placed in an electric field, will migrate towards a positive charge which is placed at the far end from the gel wells. Gel images were taken using Bio-Rad Image Lab Touch Software 1.2 on a ChemiDoc Touch Imager. Figure 6.7 and Figure 6.8 shows examples of the images obtained.

Figure 6.7. Example of an agarose gel PCR electrophoresis with intense band amplification.

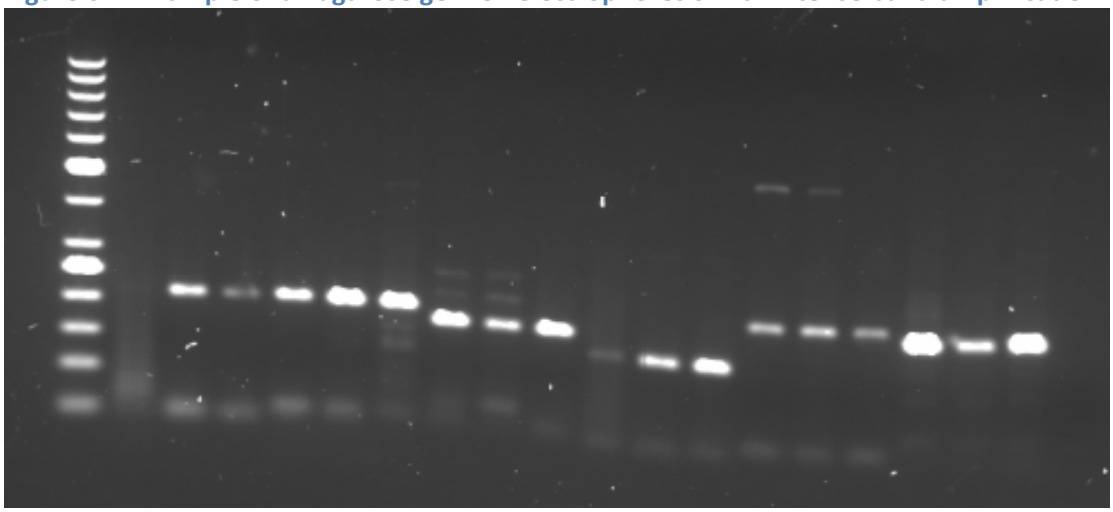
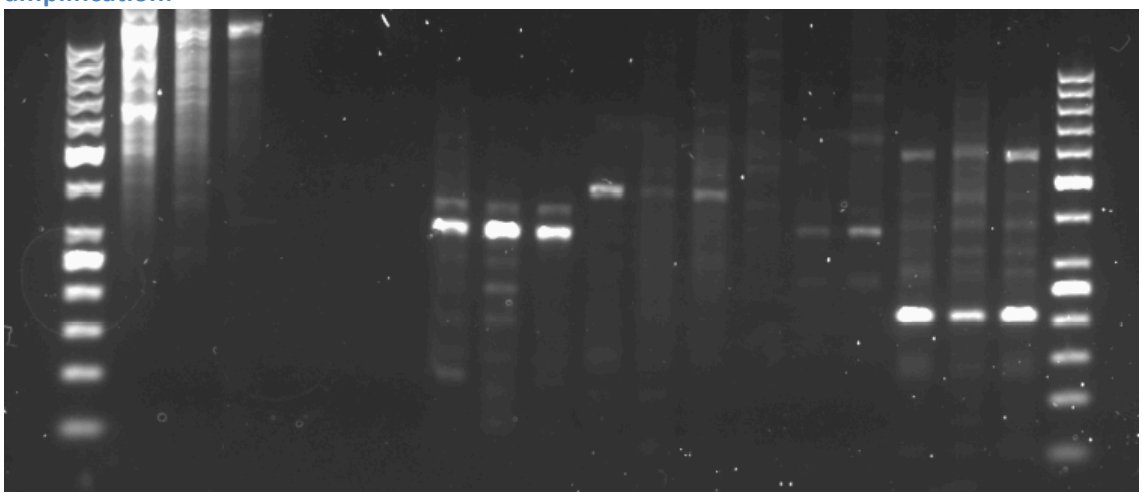


Figure 6.8 Example of an agarose gel PCR electrophoresis with non-specific band amplification.



From left to right; first column is the GeneRuler, next three columns demonstrates non specific band amplification.

6.18.2 Sanger Sequencing

Frederick Sanger and colleagues invented Sanger sequencing in 1977. It remains the gold standard in enzymatic sequencing method for variant validation. Sanger biochemistry can be applied to DNA fragments up to 1000bp in length and per base accuracy is as high as 99.999%. The principles rely on random inhibition of a PCR replicated region, creating various lengths of synthesized DNA strands. The ends of the strands are marked with fluorescent chain terminator nucleotides that allows the sequence to be determined depending on the size of the DNA strand.

The starting point of a Sanger sequencing reaction is similar to PCR, and uses PCR product as a template DNA segment. The template DNA to be sequenced is combined with primer, DNA

polymerase, four deoxynucleosidetriphosphates (dNTPs), dATP, dCTP, dGTP, dTTP and di-deoxynucleotidetriphosphates (ddNTPs) that serves as a base chain terminator by competing with its dNTP counterpart for insertion. These ddNTPs lack a 3'OH group required for the formation of a phosphodiester bond between two nucleotides, causing DNA polymerase to cease extension.

The mixture is first heated to denature and separate the double stranded DNA strands, then cooled so the primer can bind to the single stranded DNA. Next the temperature is raised again to allow amplification of new DNA using DNA polymerase. DNA polymerase will continue to add dNTPs to the chain until it happens to add a ddNTPs. Because there are many identical copies of the PCR template, a range of different fragments is produced depending on the terminal position of the dNTPs where the corresponding ddNTP has been inserted. This process is repeated for a number of cycles.

Next, laser excitation of four fluorescent labels of fragments of different lengths provides the readout in a Sanger sequencing trace. Four fluorescent dyes are used as labels for the base-specific reactions. Samples containing mixtures of all four base sequencing reactions are size-fractionated by polyacrylamide gel electrophoresis. While the fragments are migrating during the electrophoresis run, a laser beam is focused at a specific constant position on the gel. As the DNA fragments migrate past this position, the laser causes the dyes to fluoresce. Maximum fluorescence occurs at different wavelengths for the four dyes; the information is recorded electronically and the interpreted sequence is stored in a computer database.

6.18.2.1 Sanger Sequencing protocol

Clean up of PCR products

- PCR product was diluted in nuclease free water based on its intensity on the gel. For a medium intensity PCR product, this was diluted to 1:3 and for a low intensity PCR product, this was diluted either to 1:2 or was undiluted.
- Using a deep well plate, add and mix 6µl of diluted PCR product with 6µl of microclean (Webscientific limited 2MCL10). Microclean concentrates and purifies PCR product by removing reaction buffers, enzymes, primers and dNTPs.
- Centrifuge at 3500rpm for 40 minutes at room temperature.
- Remove lid from plate, invert and place paper towel underneath plate.
- Centrifuge at 1000rpm for 30 seconds.

Sequencing extension PCR

- A sequencing premix containing BigDye V1.1 is used. ABI Prism BigDye Terminator V1.1 (part number 4337450) contains sequenase enzyme, dNTPs, di-deoxy dNTPs and $MgCl_2$.
- To make up BigDye sequencing premix:

BigDye sequencing premix	**1x**
BigDye V1.1	0.5µl
Double distilled H_2O	0.5µl
5 x Sequencing Buffer (part number 4337450)	2µl

- The BigDye sequencing premix is then mixed with a forward or reverse primer and double distilled H_2O for the sequence extension PCR
- To make up sequence extension PCR mix:

Sequence extension PCR mix	**1x**
BigDye sequencing premix	3µl
Forward or reverse primer (10pmol/µl)	0.2µl
Double distilled H_2O	6.8µl

- 10µl of the sequence extension PCR mix is then added to the cleaned up PCR product and then inserted into a thermocycler programmed with cycling conditions as shown in Table 6.5.

Table 6.5 Cycling conditions for sequencing extension PCR.

Step	**Description**	**Temp (^{o}C)**	**Time**
1	Initial denaturation	96	1 minute
2	Denaturation	96	30 seconds
3	Annealing	50	15 seconds
4	Extension	60	4 minutes
5	Go to step 2 and repeat for 28 cycles		
6	Incubation	20	2 minutes

Precipitation of Sequencing Products

- After the sequence extension PCR, products are precipitated with Ethanol and 125mM EDTA to concentrate the product.

- To make up Ethanol/EDTA premix:

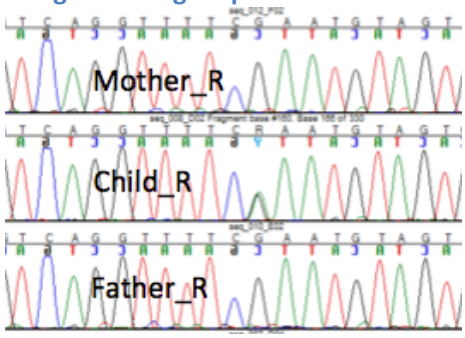
Ethanol/EDTA premix	1x
100% Ethanol	30µl
125mM EDTA	2.5µl

- 32.5µl of the Ethanol/EDTA premix was added to each sequencing product.
- Centrifuge for 3500rpm for 40 minutes at room temperature.
- Remove lid from plate, invert and place paper towel underneath plate.
- Centrifuge at 1000rpm for 30 seconds at room temperature.
- Add 50µl of freshly made 70% ethanol to each well for the final purification step.
- Centrifuge at 1650rpm for 20 minutes at room temperature.
- Remove lid from plate, invert and place paper towel underneath plate.
- Centrifuge at 1000rpm for 30 seconds at room temperature.
- Add 10µl of Hi-Di Formamide into each well that make up the sequencing injections.
- Heat for 2 minutes at 95^{o}C.
- Load onto the ABI 3130.

Review of chromatograms

The ABI 3130 produces sequencing outputs as ABI and SCF sequence. This was imported along with the reference sequences into Sequencher 5.4.1., which was used to visualise Sanger sequencing outputs. Sequence trimming and assembly was used to assemble DNA fragments to reference sequences into chromatograms to visualise and detect the variant of interest. See Figure 6.9.

Figure 6.9 Example of chromatogram using Sequencher.



R denotes sequences obtained by reverse primer, the child carries a *de novo ASH1L* G>A variant.

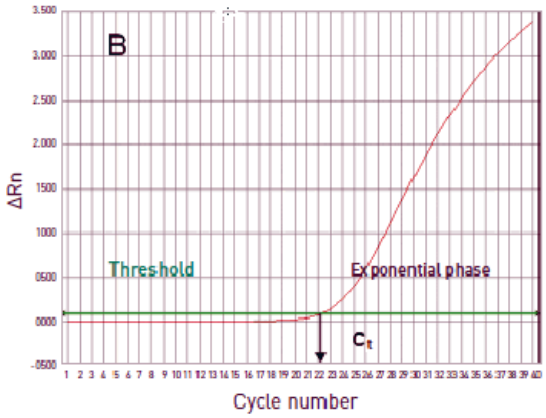
6.18.3 Real Time Polymerase Chain Reaction

Real time polymerase chain reaction (RTPCR) or quantitative PCR (qPCR) uses a fluorescence detecting thermocycler to amplify specific nucleic acid sequences and simultaneously measure their concentrations. The reaction products are analysed in an early stage in the amplification process when the reaction is still exponential, allowing more precise quantitation than at the end of the reaction[179]. The QIAGEN Type-it CNV SYBR Green PCR +qC kit (Hilden, Germany) was used for relative quantification of gene copy number and CNV validation. For each sample, a ratio value for the GOI is calculated allowing comparison of dosage to the others in the assay.

Basis of SYBR Green based qPCR Method

Quantification in RTPCR depends on detecting a fluorescence signal that is generated after binding of SYBR Green to amplified product at the early stages of the PCR reaction. SYBR Green dye shows relatively little fluorescence when free in solution but when it binds to double-stranded DNA, its fluorescence increases more than 1000 fold. This produces sigmoidal-shaped amplification plots in which fluorescence is plotted against the number of cycles. The threshold cycle (CT) value is obtained from the plot. This is the point at which there is a first significant detectable increase in fluorescence on the linear point of the amplification curve and is used for calculation of the starting template amount in each sample. Figure 6.10 shows a graphical representation of real time PCR data.

Figure 6.10 Graphical representation of real time PCR data.



Rn fluorescence of the reporter dye divided by the fluorescence of a passive reference dye is plotted against PCR cycle number. CT is the intersection between an amplification curve and threshold. Figure obtained from http://www.thermofisher.com.

The comparative or ΔΔCT threshold point relies on direct comparison of CT values of the target sample GOI and a control sample/s with a reference assay. The Type-it CNV reference primer assay is a universal reference assay and relative quantification control. This allows

amplification of a multi-copy genetic element that can be used as a reference for reliable ΔΔCT based relative copy number quantification analysis. The copy number of the reference gene is not changed among different samples; therefore any CT difference of the reference assay reflects the difference of the template DNA amount in the target sample.

Primers for the gene of interest (GOI) were designed as described previously. Two sets of primers for each CNV region were designed as described previously (see section 6.19.1.1); one covering the start of the exon and the other about two-thirds into the region (Primers used were named KLHL1_1 and KLHL1_2). Primers were diluted to 17.5µmol.

A reference and sample GOI mastermix was prepared as follows:

Reference assay mastermix 1x	**Sample GOI mastermix 1x**
5µl SYBG Green mastermix 0.4µl reference primer	5µl SYBG Green mastermix 0.4µl GOI primer (0.2ul forward and 0.2ul reverse primer)

The Type-it SYBR Green PCR master mix contains HotStarTaq plus DNA Polymerase, Type-it SYBR Green PCR Buffer and dNTPs. Each reaction is done in triplicate and the mean calculated. 5µl DNA (2ng/µl) was mixed with 5.4µl of the mastermix and the plate loaded into the ABI PRISM 7900 system (Applied Biosystems, Foster City, CA), which was programmed settings as shown in Table 6.6:

Table 6.6 Cycling conditions for real time PCR.

Step	**Time**	**Temperature**
Initial PCR activation step	5 minutes	95°C
2-step cycling Denaturation Annealing/extension	 30 seconds 30 seconds	 95°C 60°C
Number of cycles	35	

The real-time cycler is set up for automatic calculation of threshold and baseline although these default values are checked to make sure that the threshold is above the background and significantly below the plateau of an amplification plot on the linear region of the amplification curve.

The PCR efficiency of each primer pair was checked over a dilution series of DNA (30ng, 3ng, 0.3ng and 30pg) for comparability with the proprietary reference assay of a multicopy gene.

Mean CT amounts were calculated and plotted against Log DNA concentrations for reference DNA 1050 with reference, KLHL1_1 and KLHL1_2 primers. TE Buffer was used as a No Template Control.

When the reaction is complete, the following calculations are performed: The copy number change (CN) for the sample is calculated by the differences of ΔCT between the GOI and reference assay and then calculating the ΔΔCT values between the GOI and control.

$$\Delta CT = CT(GOI) - CT(reference)$$

This is the difference between the CT value of the GOI (KLHL1_1 or KLHL1_2) and the CT value of the reference assay (reference gene and reference DNA 1052 or 1039)

The ΔΔCT value is then calculated. This is the difference between the mean ΔCT value of the sample (e.g. 00518) and the mean ΔCT value of the reference (1052 or 1039).

$$\Delta\Delta CT = mean\Delta CT(sample) - mean\Delta CT(reference)$$

The copy number change is calculated for the sample relative to the control by the formula:

$$CN = 2^{\wedge -\Delta\Delta CT}$$

If CN>1, the copy number of the GOI is higher in the test sample than the control sample;
if CN<1, the copy number of the GOI is lower in the test sample than the control sample.

Chapter 7 Results: Genetics

7.1 Sequencing cohorts

The exome sequencing cohort (n=109) comprises cases within the same three phenotyping groups; the UK cohort (n=49), the Euroepinomics cohort (n=42) and the Italian cohort (n=18). I recruited the UK cohort (see section 2.2), the Euroepinomics cohort was assembled by European collaborators (see section 2.3.2), and the Italian cohort was provided by Professor Renzo Guerrini and Professor Carla Marini (Anna Meyer Children's Hospital, Florence, Italy). The UK and Italian cohorts were sequenced at the Guy's Genomics Facility while the Euroepinomics cohort consisted of 40 European trios sequenced at Wellcome Trust Sanger Institute and two cases from a multiplex family sequenced at the Danish Epilepsy Centre in Denmark. Out of the 109 cases with MAE in the exome sequencing cohort, three cases had family studies performed where relevant family members were also exome sequenced and analysed. All exome sequencing data was aligned, variant called and annotated using the Guy's Genomic Facility pipeline.

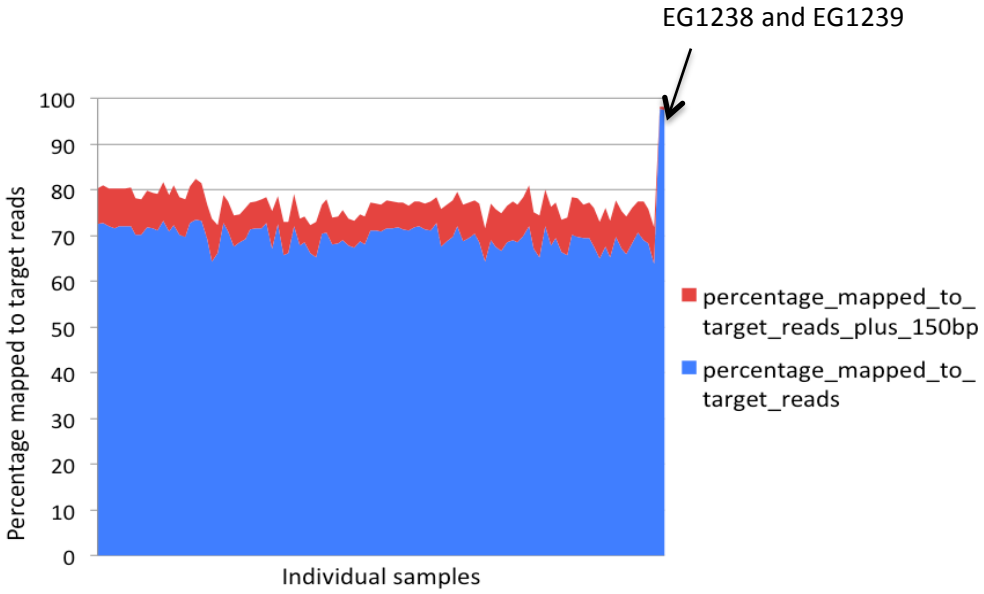
Sample ID and genders of cases collected for the UK cohort and Italian cohort is presented in appendix G. An exome sequencing ID is available when this was performed. DNA collected from family members in the UK cohort is also indicated.

7.2 Summary of QC metrics in exome sequencing cohort

7.2.1 Capture efficiency and read depth

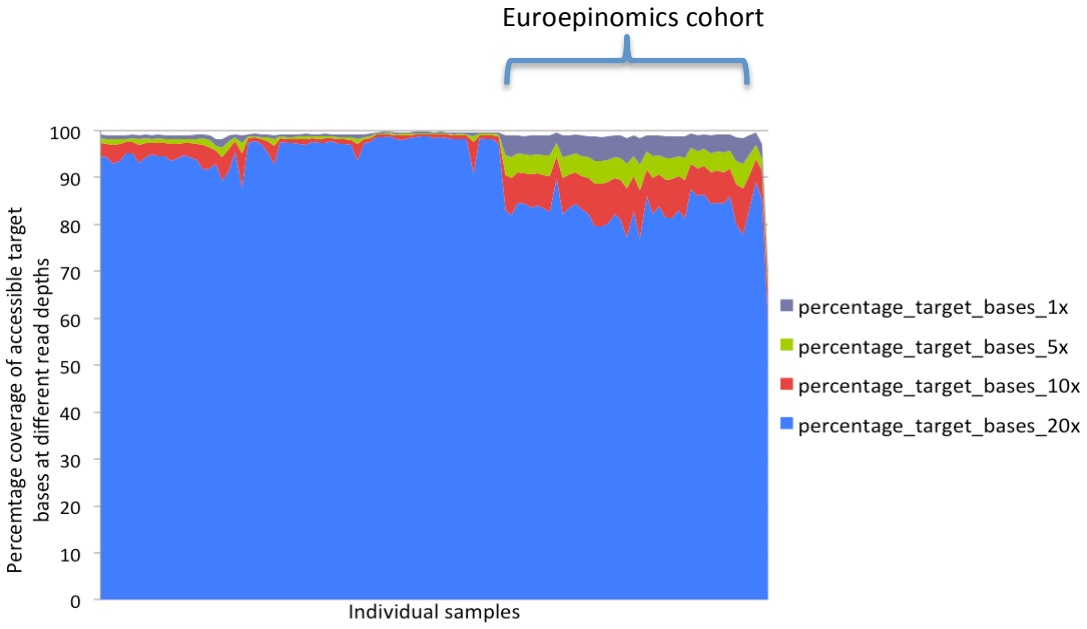
The mean percentage of reads mapped to target regions was 70.07% (SD 4.51), and when expanded to reads within 150bp of target, 77.31% (SD 3.79) was achieved (see Figure 7.1). Target reads are reads that align to target regions, which are exome regions that are selected and captured by specific biotinylated probes during the exome sequencing process. The two samples from the Danish Epilepsy Centre EG1238 and EG1239 were clear outliers with 97.64% and 97.57% mapped to target reads and 98.28% and 98.22% within 150bp respectively. When these two samples are removed from the cohort, the mean falls to 69.54% (SD 2.37) mapped to target reads and 76.90% (SD 2.44) mapped to target reads plus 150bp.

Figure 7.1. Capture efficiency to target reads.



The mean percentage of coverage of accessible target bases (bases within target reads) at read depth of 20x was 90.5% (SD 7.49), although there was a clear difference between samples sequenced at the Guy's Genomics Facility and elsewhere (see Figure 7.2). When analysed separately, samples sequenced at the Guy's Genomics Facility achieved 95.5% (SD 2.60) coverage of 20x compared with the Euroepinomics cohort, which achieved 83.0% (SD 2.84). Sample EG1238 and EG1239 from the Danish Epilepsy centre achieved 85.16% and 59.26% coverage of 20x respectively.

Figure 7.2. Read depth of accessible target bases at 1x, 5x, 10x and 20x.

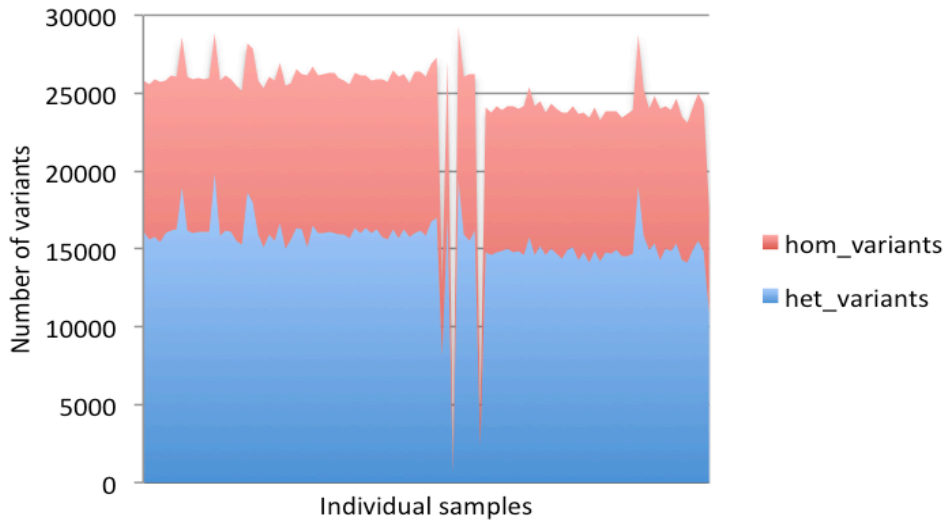


No samples were rejected as this stage. This was because it was felt that errors resulting from reduced capture efficiency or read depth would mainly result in false negatives. In addition these QC errors could be filtered out at individual sample and variant call level by direct interrogation using IGV.

7.2.2 Variant characteristics

Annotated variant characteristics and variant numbers in each sample were reviewed. Per sample, the mean total number of coding variants was 25333 (SD 1537), there was a mean of 15656 (SD 1185) heterozygous variants and 9677 (SD 549) homozygous variants (see Figure 7.3). Three samples were removed from further analysis as they had insufficient variants called. They were S2539 with 957, S2544 with 3930, and S2537 with 12747, total numbers of variants. Given the importance of novelty in filtering analysis, the patterns of novel variants were explored in the remaining samples.

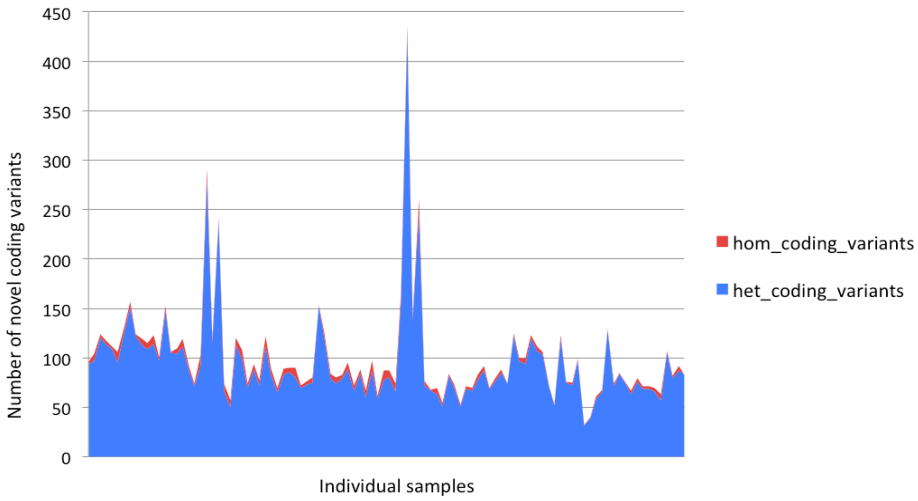
Figure 7.3 Number of variants.



hom_variant homozygous variants, het_variants heterozygous variants, troughs in graph from left to right represent variant counts from S2537, S2539 and S2544.

In the samples remaining after the three samples with insufficient variants were removed, there was a mean of 104 (SD 52.1) novel coding variants consisting of a mean of 99.9 (SD 51.1) novel heterozygous variants and 4.5 (SD 2.8) homozygous novel variants (see Figure 7.4). The three samples with the most novel variants were ethnically diverse; S2536 of Asian Gujarati descent with 440 novel variants, S2286 with 300 and S2540 with 274 *novel* variants, both of Black African descent. This is probably because control population variant databases do not sufficiency account for these ethnic distribution compared to white Caucasian populations.

Figure 7.4. Number of novel heterozygous and homozygous coding variants.



hom_coding homozygous coding variants, het_coding heterozygous coding variants, highest four peaks in graph from left to right are S2286, S2341, S2536 and S2540.

7.2.3 Transition – Transversion Ratio

The transition – transversion ratio is a useful diagnostic metric. Transitions involve bases of similar shape and are interchanges of purine (A <-> G) or of pyrimidine (C <-> T). Transversions involve exchange of purine for pyrimdine bases (A <-> C, A <-> T, G <-> C or G <-> T). Transitions occur more frequently than transversions due to shifts during replication. In exome sequencing, we would expect a ratio of about 3.0. The mean transition-transversion ratio in the MAE exome sequencing cohort was 2.96 (SD 0.04).

7.3 Annotation analysis of reported epilepsy associated gene variants

7.3.1 All_epilepsy variants

3254 unique All_epilepsy variants were assembled from ClinVar, HGMD and EpilepsyGene databases. The genetic variants curated in this set consisted of disease cohorts associated with epilepsy such as metabolic disorders, intellectual disability syndromes, congenital syndromes as well as epilepsy syndromes.

7.3.1.1 Minor Allele Frequency

Annotations of variants with MAF from 1000 Genomes, ESP and ExAC revealed that even amongst this heterogenous cohort, variants were mostly novel or ultra rare (MAF <0.0001). Weighted MAF calculations demonstrated that 83.7% of all variants were novel, 99.8% of all

variants had a MAF of <1% and 98.5% of all variants had a MAF <0.1% (Table 7.1). Additionally further review of the 6 variants with the highest MAF variants, with weighted MAF (0.1 to 0.99) showed that they were annotated insufficiently, benign in three cases and in one case, a risk factor for familial febrile seizures (see Table 7.2).

Table 7.1. Distribution of weighted MAF of All_epilepsy variants.

Weighted MAF	Number of variants (% of total)
0.1 to 0.99	6 (0.2%)
0.01 to 0.099	19 (0.6%)
0.001 to 0.0099	46 (1.4%)
0.0001 to 0.0009	152 (4.6%)
0.00001 to 0.000099	200 (6.1%)
0.000001 to 0.0000099	105 (3.2%)
0 (Novel)	2726 (83.7%)

Table 7.2. Clinical significance of All_epilepsy variants with highest weighted MAF.

Gene	Chr:position ref>alt	dbSNP ID	Weighted MAF	Clinical significance
SCN1A	2:166909544 C>T	rs3812718	0.4934	Risk factor for familial febrile seizures
NHLRC1	6:18122506 G>A	rs10949483	0.6196	Benign
CHRNA7	15:32460616 C>T	rs2229956	0.2857	No information
CACNA1A	19: 13397560 C>T	rs16027	0.1578	Benign/Likely benign
GJD2	15:35045057 G>A	rs3743123	0.5223	In HGMD
SCN1B	19:35524944 G>C	rs67486287	0.2124	Benign

7.3.1.2 *In silico* predictors

In silico prediction of the All_epilepsy variants with SIFT, PP2hvar and CADD was performed. Individually, SIFT predicted pathogenicity in 64.8% (1294 out of 1995), PP2hvar in 74.0% (1261 out of 1704) and a CADD score of 20 captured 86.0% (2025 out of 2353) of variants. When used together, the three predictors annotated 51.2% of All_epilepsy variants pathogenic (Figure 7.5). CADD appeared to be the best performing predictor and at a lower cut off score of 10 (top 1% most deleterious) captured 96.3% of variants (see Figure 7.6).

Figure 7.5. Venn diagram of All_epilepsy variants annotated pathogenic by SIFT, PP2hvar and CADD.

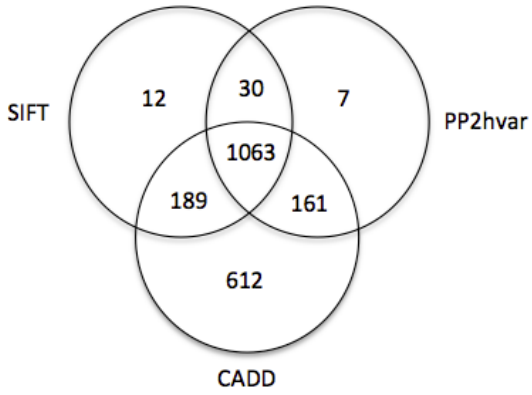
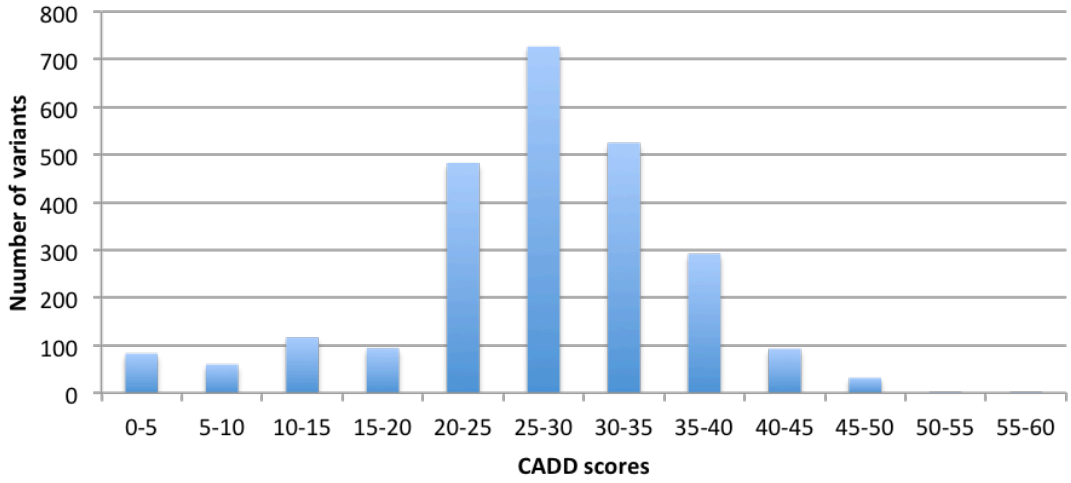
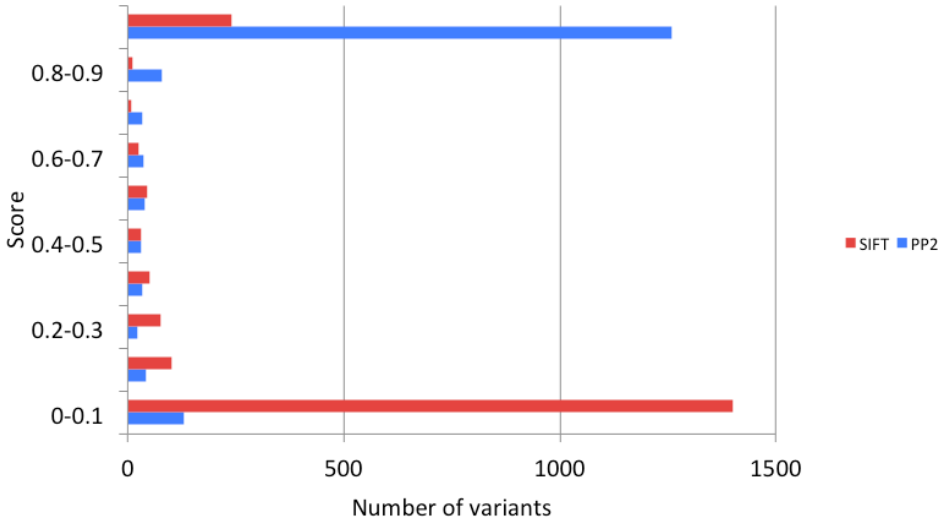


Figure 7.6. Distribution of CADD scores for All_epilepsy variants.



As an internal validation, the inverse relationship between SIFT and PP2hvar scores was also explored. The pathogenic cut off scores for SIFT and PP2hvar are opposite; SIFT considers scores <0.05 as pathogenic whereas PP2hvar considers scores >0.909 pathogenic. Therefore for any given presumed pathogenic variant, it should have a low SIFT score and high PP2hvar score and vice versa for a benign variant. This expected inverse relationship in SIFT and PP2hvar scores due to their properties as described is shown in Figure 7.7.

Figure 7.7. Inverse relationship between SIFT and PP2hvar scores of All_epilepsy variants.



7.3.2 Pure_epilepsy variants

Pure_epilepsy autosomal dominant

622 unique variants from 45 genes made up the Pure_epilepsy autosomal dominant group following stratification of the ClinVar search results. 97.6% of variants were novel with the remaining 15 variants with a median weighted MAF of 9.03×10^{-5} (range 8.2×10^{-6} to 3.17×10^{-3}). As expected, *in silico* predictors predicted more variants to be pathogenic in this group compared to the All_epilepsy group since these variants represented specific epilepsy syndrome phenotypes. SIFT predicted pathogenicity in 82.4% (320 out of 388), PP2hvar 80.4% (305 out of 379) and a CADD score of 20 and above captured 93.0% (400 out of 430) of variants. When SIFT, PP2hvar and CADD are used together, this captured 64.8% (279 out of 430) variants.

Pure_epilepsy autosomal recessive

From the Pure_epilepsy group, 41 variants were stratified as having an autosomal recessive inheritance following individual variant review of the ClinVar search results. Of these only 3 variants were annotated (Table 7.3).

Table 7.3. Pure_epilepsy autosomal recessive variant annotation.

Annotations	11: 792735 C>G	11: 749509 A>G	19: 2438464 T>C
1000 Genomes MAF	0.00019		
ESP MAF	0.0006		
ExAC MAF	0.0004		
CADD score	-	27.6	25.4
SIFT score		0.07	0.04
PP2hvar score		0.996	0.932

7.4 De novo analysis of Euroepinomics MAE trios

The Euroepinomics consortium sequenced and analysed 40 MAE trios for *de novo* variants. Nine pathogenic *de novo* mutations from nine different genes were identified, six of which have been published (*CHD2*[15], *SYNGAP1*[16], *KCNA2*[19], *STX1B*[20], *SMARCA2*[52], *KIAA2022*[23]).

Table 7.4 summarises the variant characteristics and clinical phenotypes of these cases. The most striking features in these solved cases are the associated neurodevelopmental problems. Eight out of nine cases had evidence of ID and six cases had psychiatric comorbidity in the form of ASD, ADHD, obsessive-compulsive disorder or aggressive behaviour.

Table 7.4. Pathogenic *de novo* mutations identified in the Euroepinomics exome sequencing trios.

Subject ID	Gene	c.DNA: protein change	SIFT, PP2, CADD	gnomAD MAF	Seizure phenotype	EEG	Additional clinical features
[15]IP09MAE01p W_1145084	*CHD2*	c.4771delC:p.L1591X	-, -, -	Novel	12y M, seizure onset at 3.5y with myoclonic, MA, GTCS and absence seizures.	Abnormal	Moderate ID, ADHD and aggressive behaviour
[19]8MAE23 W_1152620	*KCNA2*	c.T788C:p.I263T	0 (P), 0.977 (P), 22.7	Novel	7y M, seizure onset at 11m with myoclonic and MA. Seizure free since 4y.	Multifocal sharp waves and spikes	Prior DD, mild-moderate ID
[23]EG1263 W_1144306	*KIAA2022*	c.C964T:p.R322X	1 (T), -, 37	Novel	11y F, seizure onset at 30m with myoclonic, atonic, tonic and focal seizures. Ongoing seizures.	Polyspike wave and focal discharges	Mild ID, ASD, ADHD
IP09MAE02p W_1158858	*NBEA*	c.7230delA:p.P2410fs	-, -, -	Novel	4y M, seizure onset at 19m with myoclonic, GTCS, absences seizures	Generalised epileptic activity	Mild ID, ASD, ADHD, aggressive behaviour
EG1268 W_1145970	*SCN1A*	c.C4540T:p.R1514X	1 (T), -, 45	Novel	14y F, seizure onset at 3m with GTCS, atonic, myoclonic and absence seizures	Generalised spike and slow waves	Mild ID, ASD and OCD like behaviour
9329/W_1145223	*SCN2A*	c.C2790A:p.H930Q	0 (P), 0.999 (P), 27.2	Novel	4y M, seizure onset at 17m with atonic seizures, GTCS	Generalised epileptic activity	ASD, ADHD
[52]3003301 S1272	*SMARCA2*	c.C3721G:p.Q1241E	0.5 (T), 0.954 (P), 22.5	Novel	5y F, seizure onset at 14m with MA and GTCS. Seizure free since 4y.	Generalised spike wave and polyspike	Severe ID, ASD, severe feeding difficulties, dysmorphism
[20]18_P W_1199460	*STX1B*	c.G676C:p.G226R	0 (P), 1 (P), 32	Novel	6y F, seizure onset at 13m with GTCS, MA, myoclonic and tonic seizures	Generalised epileptic activity	Prior febrile convulsions Moderate ID. Hypotonic with unsteady gait
[16]4MAE10 W_1158897	*SYNGAP1*	c.T1995A:p.Y665X	0.85 (T), -, 35	Novel	12y M, seizure onset at 1y with atonic and MA seizures. Ongoing seizures.	GSW, photosensitive	Severe ID with absent speech. Unsteady gait

P pathogenic, T tolerated, y year, m months, M male, F female, MA myoclonic atonic seizures, GTCS generalised tonic clonic seizures, DD developmental delay, ID intellectual disability, ADHD attention deficit hyperactivity disorder, ASD autism spectrum disorder, OCD obsessive compulsive disorder, GSW generalised spike wave.

Two genes that were not previously associated with MAE were identified by other members in the Euroepinomics consortium; *NBEA* and *SMARCA2.*

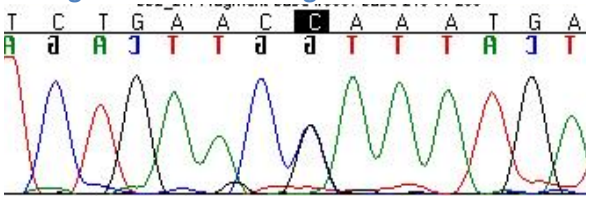
NBEA

NBEA (Neurobeachin) encodes an A kinase anchoring protein that was identified as a regulator of membrane protein trafficking and is required for the formation and functioning of central synapses[180]. *NBEA* is not a previously recognised epilepsy gene and this *de novo* variant is yet unpublished. Further cases with pathogenic variants have been collected and functional work on these variants is being carried out by other investigators (personal communication, Euroepinomics MAE consortium).

SMARCA2

SMARCA2 encodes the catalytic subunit components of the SWItch/sucrose nonfermentable like chromatin remodelling complex (SWI/SNF complex). Variants in the SNF2 ATPase domain of *SMARCA2* are associated with 80% of patients with Nicolaides Baraitser syndrome[181]. This was a London case and deep phenotyping and Sanger sequencing was carried out here. Subject 3003301 had a *de novo SMARCA2* missense mutation c.3721C>G,p.Q1241E (ENST00000349721 NM_0030703). Sanger sequencing confirmed this in the patient (Figure 7.8). This variant is novel on gnomAD and there are only very rare missense and stop gain variants in the region (MAF< 1 x 10^{-5}).

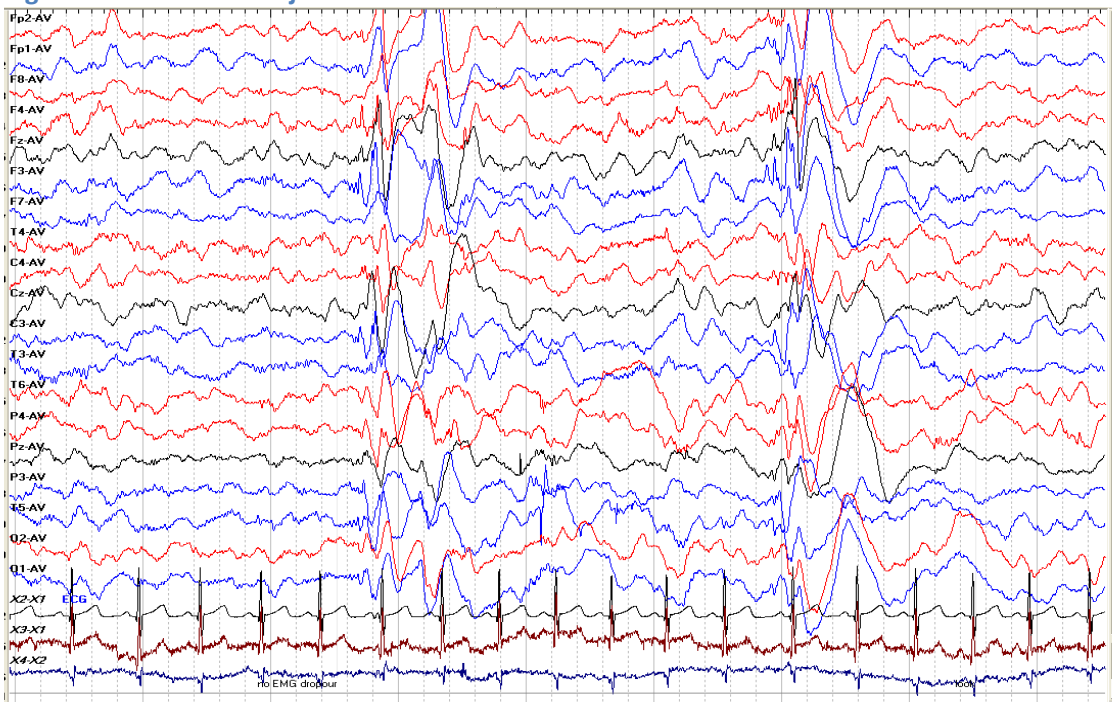
Figure 7.8. Chromatogram demonstrating *SMARCA2* substitution in subject 3003301.

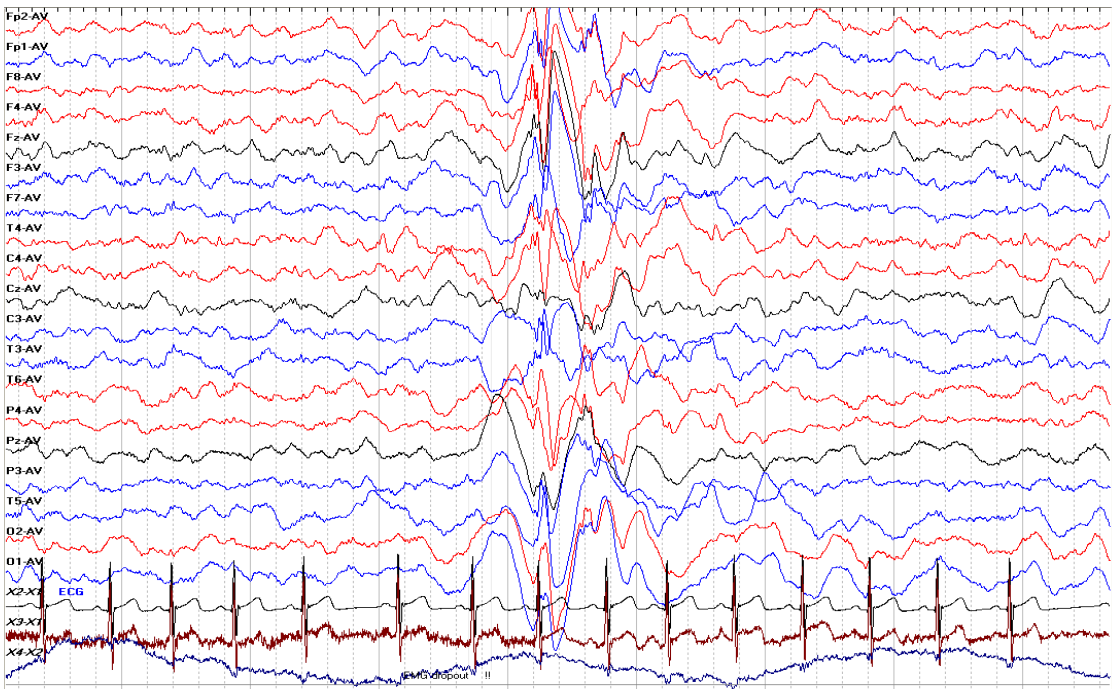


Subject 3003301 is a girl of Ghanaian parentage, was born at term with a birth weight of 3.4kg (50th centile) and a head circumference of 32cm (2nd centile). Her mother referred normal early developmental milestones as she sat unsupported at 8 months, walked at 12 months and was babbling just before her first year. However, her mother had concerns that she had reduced visual interest in people and did not smile readily. Her first seizure was at 14 months and was a myoclonic atonic seizure, which continued as her prominent seizure type while she developed generalized tonic-clonic and absence seizures. At 17 months her weight was 10.7kg (50 - 75th centile), height 82.5cm (75 - 91st centile) and head circumference 48cm (98 – 99.6th centile). Her EEG demonstrated frequent generalized bursts of polyspike and wave activity

during wakefulness and sleep against a normal background. Electrographic correlate of negative axial myoclonus was captured (see Figure 7.9). An MRI brain scan was normal. Array CGH analysis with ~ 44 000 probes was normal. Developmental assessment showed a delay in language and social communication skills. Additionally, the patient had feeding problems from 6 months following the introduction of solids. She was hypersensitive to textures and demonstrated food refusal and would hold food in her mouth or vomit during meal times. No dysmorphic features were identified at the time. Based on these electoclinical features, her epilepsy type was thought compatible with a diagnosis of MAE syndrome. She was treated with sodium valproate to which she responded, attaining complete seizure remission at 3 years with normalization of her EEG. At 2 years 9 months her weight was 13.9kg (25 - 50th centile), height 94.7cm (50 – 75th centile) and head circumference was 49cm (9 - 25th centile). Antiepileptic medication was discontinued at 4 years. At 5 years, she had developed a restricted pattern of feeding with aversion to lumpy foods. The 3di autism interview revealed abnormalities in communication and non-verbal communication, with borderline scores in social reciprocity and restricted/repetitive behavior and interests and she was diagnosed with autism spectrum disorder. On neurodevelopmental assessment with the WPPSI III, she demonstrated severe learning difficulties and scored 51 for verbal IQ (0.1th centile), 57 for performance IQ (0.2th centile) and 47 for Full Scale IQ (<0.1th centile). Repeat awake and sleep EEG and repeat MRI brain scans were normal.

Figure 7.9. EEGs for subject 3003301.





The EEGs shows generalised spike and polyspike wave against a normal background (top); and negative myoclonus associated with an electrodecrement (bottom).

Clinical examination at 5 years 5 months showed skin wrinkling, frontal balding, broad and long philtrum, broad nasal base, upturned nasal tip, thick alae nasi, anterior projection of the upper lip over the premaxilla, thin upper vermillion and thick lower vermillion border. She did not have brachydactyly or prominent interphalangeal joints (see Figure 7.10). She had eczema particularly involving the elbows and knees. Her clinical phenotype of specific craniofacial features, reduced speech, ID, ASD and seizures along with a *de novo SMARCA2* mutation were felt to be consistent with the Nicolaides-Baraitser syndrome spectrum[182].

Figure 7.10 Clinical photographs demonstrating features of Nicolaides Baraitser syndrome.

(a) and (b) Skin wrinkling, frontal balding, broad nasal base, upturned nasal tip, thick alae nasi, broad and long philtrum, thin upper vermillion, thick lower vermillion, anterior projection of the upper lip over the premaxilla. (c) Clinical photographs of hands with slightly thick distal phalanges. (d) Clinical photograph of feet.

7.4.1.1 Exploration of the role of SMARCA2 in other MAE cases

Variants in *SMARCA2* were reviewed in the rest of the MAE exome sequencing cohort. Two cases had novel/rare *SMARCA2* variants.

WTSI_MAE_1158850 had a c.1746+7 G>T variant inherited from an unaffected father. Splice site predictors (Ada 0.0.0002, RF 0.016) were not significant, gnomAD MAF novel. Subject WTSI_MAE_1158850 is a 3 years 8 months old male with previous normal development, he developed myoclonic atonic seizures at 3 years. At the time of recruitment, he had no associated neurodevelopmental comorbidity. Euroepinomics collaborators from Bucharest performed neurological examination and EEG with photosensitive stimulation on the father, which was normal (personal communication Dr. Dana Craiu). The variant was classified as inconclusive during IGV interrogation due to a read depth of 11 and 17 reads in the child and father. The variant is located in exon 10, away from SNF2 ATPase domain. The variant was classified as a VUS.

WTSI_MAE_1153136 has a c.A337G,p.M113V variation with variant prediction scores of SIFT 0.39 (tolerated), PP2hvar 0.339 (tolerated), CADD 25.4. The variant is inherited from his unaffected mother. Subject WTSI_MAE_1153136 is a 17-year-old male from Kiel. He had previous normal development and had his first seizure at 3 years, which was a myoclonic-atonic seizure, which remained his only seizure type. He went into seizure remission at 5 years. He has no associated neurodevelopmental difficulties. There is a family history of epilepsy in the paternal cousin with Jeavon's syndrome. (personal communication Dr. Sarah von Spiczak). This variant has a MAF of 3.54×10^{-5} on gnomAD and is located on exon 3, away from the SNF2 ATPase domain. This variant was classified as a VUS.

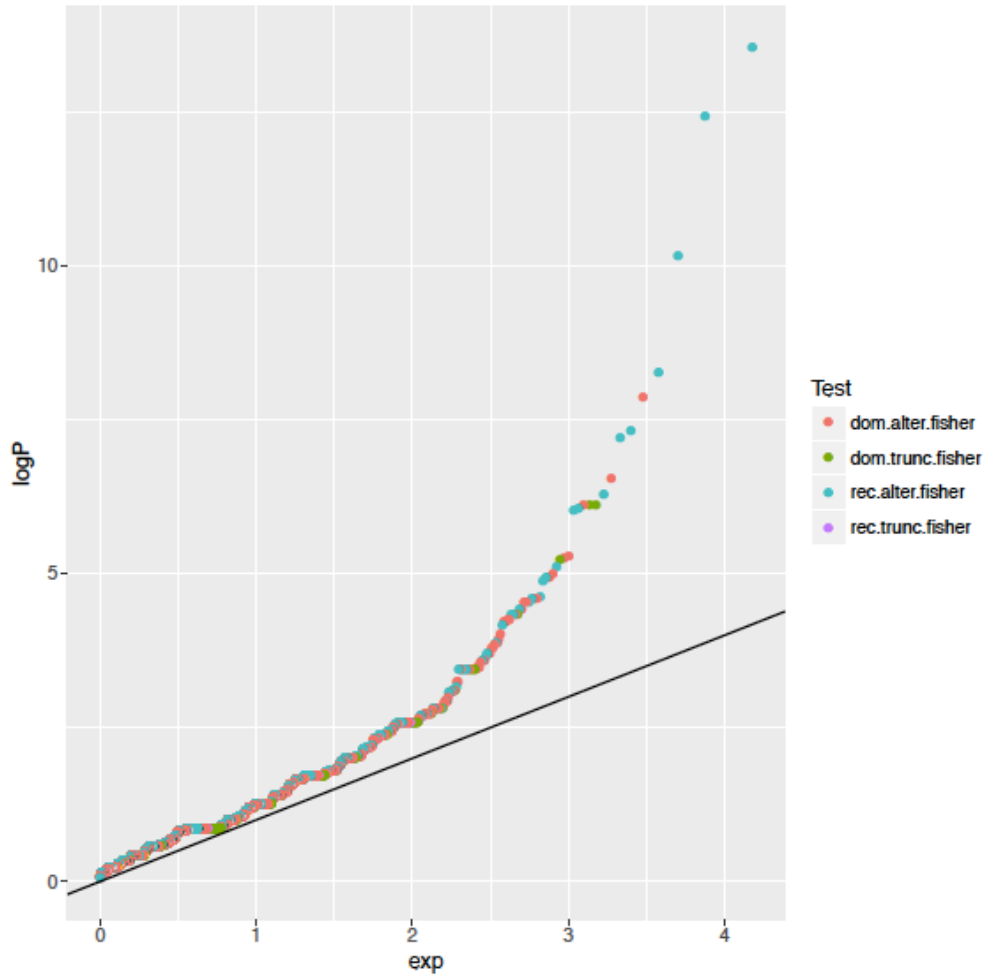
7.5 Case control association analysis

Case control association analysis of 52 cases and 524 ancestry-matched controls were tested for enrichment of rare variants with a Fisher test using EPACTS v3.2.3. Four models were tested; dominant (MAF<0.01%) or recessive (MAF<0.05%), and then altering or truncating variants. The top ten genes with the lowest *P* values in the four models tested are listed on Table 7.5. The distribution of observed and expected $\log_{10}$ *P* values is shown on the quantile-quantile plot in Figure 7.11. Literature reviews were carried out on the genes but it was not possible to pick out an obvious candidate. This was one of the first gene identification strategies that was attempted and it was subsequently abolished mainly due to the understanding that most epilepsy/MAE variants discovered to date were novel and this model was more suited for testing rare variants; eliciting genes that were enriched with rare variation compared to controls. Additionally the extent of the genetic heterogeneity of MAE was emerging following the success of NGS collaborative studies, and the small sample size of this cohort would also limit the power of discovery.

Table 7.5 Genes with smallest *P* values in case control association analysis.

Dom.Alter		Dom. Trunc		Rec.Alter		Rec.Trunc	
Gene	***P* value**	**Gene**	***P* value**	**Gene**	***P* value**	**Gene**	***P* value**
RBMS1	1.37×10^{-8}	*STRAP*	7.65×10^{-7}	*SH2D2A*	2.79×10^{-14}	*IPO5*	0.002649
UTP23	2.85×10^{-7}	*TFAP2D*	7.65×10^{-7}	*PHF21A*	3.69×10^{-13}	*PARK2*	0.002649
MAGOHB	7.65×10^{-7}	*KIAA1432*	6.03×10^{-6}	*GOLGA6L2*	7.02×10^{-11}	*AF131215.4*	0.002649
KRT16	5.07×10^{-6}	*GRIK2*	4.66×10^{-5}	*GXYLT1*	5.31×10^{-9}	*SSX9*	0.002649
ACADSB	5.45×10^{-6}	*SMC3*	0.000354	*TAS2R46*	4.78×10^{-8}	*DSTYK*	0.004625
PARD6A	1.00×10^{-5}	*C11orf49*	0.000354	*HNRNPC*	6.40×10^{-8}	*AC107977.3*	0.008898
UBA3	1.11×10^{-5}	*IPO5*	0.000354	*CCNYL2*	5.11×10^{-7}	*APIP*	0.009532
TNIK	2.47×10^{-5}	*KIAA2018*	0.000354	*MED15*	8.91×10^{-7}	*NUP214*	0.009532
MED15	2.51×10^{-5}	*RARS*	0.000354	*CDC27*	9.14×10^{-7}	*DDX58*	0.009532
ANKFN1	2.98×10^{-5}	*MST1P9*	0.000780	*SLC45A2*	7.57×10^{-6}	*MTCH2*	0.01594

Figure 7.11 Quantile-quantile plot of gene collapsing Fisher's exact test.



y coordinates are observed $\log_{10}$ *P* values whereas x coordinates are expected $\log_{10}$ *P* values. dom.alter dominant altering, dom.trunc dominant truncating, rec.alter recessive altering, rec. trunk recessive truncating. Qualifying variants for dominant had a MAF<0.01% and for recessive a MAF<0.05%.

7.6 Review of known epilepsy genetic variants

97 cases were reviewed for the presence of novel variants in 100 known epilepsy genes (see section 6.18 for epilepsy associated gene list). The nine cases (see section 7.4) from the Euroepinomics sequencing cohort with classified pathogenic *de novo* variants were removed from further analysis.

46 novel nonysnonymous variants/indels within 27 epilepsy genes were identified. Following IGV interrogation, 17 variants failed this step and were removed from further analysis (see Table 7.6). IGV interrogation was considered failed if (1) variant call fell in a region where total read depth was less than 10 reads, (2) variant call was in less than 20% of total reads for heterozygous mutations, (3) identical indel in more than one sample and in other in-house sample, (4) SNV more than 10bp from intro/exon boundary and (5) Indel more than 2bp from intron/exon boundary. The remaining 29 genetic variants were classified into likely benign (n=6), variant of uncertain significance (n=16) or likely pathogenic (n=7), following a detailed review of relevant literature, variant characteristics and correlation with phenotype. Three cases had more than one novel variant in an epilepsy associated gene. Subject WTSI_MAE_1195700 had variants in *KIAA2022* and *PLCB1* variant inherited from her mother, subject WTSI_MAE_1150147 had a *GRIN2A* variant inherited from her father and a *POLG* and *SYNGAP1* variant inherited from her mother; and subject S1383/768H had a variant in *CACNA1A* and *MTOR*. The MTOR variant was Sanger validated and found to be inherited from his mother.

The following section will detail the individual genetic variants described above and the reasons for their classification. Appendix I details the IGV plots for the variants described in this section.

Table 7.6. Novel epilepsy variants that failed IGV interrogation.

Subject ID	Gene	Variant	IGV analysis
S2294/00554	*ALH7A1*	chr5:125929019: ACACACACACACACACAC -	Intronic indel
S1273/3004301	*GABRG2*	c.632-6->TTTTG	Indel within 10bp of intron/exon boundary
S1377/204D	*FASN*	chr17:80050564 GGGC; GC	Same indel in multiple samples
S1385/347H	*FASN*	chr17:80050564 GGGC; GC	Same indel in multiple samples
S1387/565D	*FASN*	chr17:80050564 GGGC; GC	Same indel in multiple samples
S2341/00505	*MECP2*	c.376_377del: p.126_126del	Deletion in 2 reads only
W_M_1151131	*RELN*	chr7:103202402 CCAAAA; CAA	Multiple regional calling errors
S2286/00500	*RELN*	exon48:c.7494_7496del: p.2498_2499del	Deletion in 1 out of 4 reads
S1390/729H	*SIK1*	chr21:44846876: TGCCGCCGC; -	Same indel in multiple samples
S1273/3004301	*SIK1*	chr21:44846876: TGCCGCCGC; -	Same indel in multiple samples
S1381/262N	*SIK1*	chr21:44846876: TGCCGCCGC; -	Same indel in multiple samples
S2289/00529	*SIK1*	chr21:44846876: TGCCGCCGC; -	Same indel in multiple samples
S2332/00561	*SIK1*	chr21:44846876: TGCCGCCGC; -	Same indel in multiple samples
S2332/00561	*SLC6A1*	chr3:11057425; G>A	Total read depth 9 reads.
W_M_1128259	*SMARCA2*	c.677_678insACA: p.Q226delinsQQ	Insertion in 1 read only
W_M_1144306	*SMARCA2*	c.2349-10->T	Multiple regional calling errors
S2293/00551	*SMARCA2*	chr9:2015407; A>T	Total read depth 4 reads

7.6.1 Likely benign variants

Six different variants from six different genes were classified as likely benign. Table 7.7 summarises the details of these gene variants.

Table 7.7. Summary of likely benign epilepsy variants.

Sample ID	Gene	c.DNA: protein change	SIFT, PP2hvar, CADD or splice predictors	gnomAD MAF	Inheritance
S2294/ 00554	*ALDH7A1*	c. G1072A:p.E358K	0.34 (T), 0.009 (T), 24.4	1.98 x 10^{-5}	NK
S2341/ 00505	*HUWE1*	c.5161+9C>A	-	Novel	NK
W_M_115 0147	*KIAA2022*	c.C1418T:p.S473F	0.15 (T), 0.037 (T), 10.70	1.64 x 10^{-5}	Mother
W_M_119 5700	*PLCB1*	c.1514-4A>T	Ada 0.003 (not sig.), RF 0.058 (not sig.)	Novel	Mother
W_M_115 0147	*POLG*	c.3242_3249del: p.1081_1083del	-	-	Mother
S1381/ 262N	*TBC1D24*	c.T686C:p.F229S	0.01 (P), 0.97 (P), 24.3	Novel	NK

T tolerated, P pathogenic, sig significant, NK not known

ALDH7A1

The *ALDH7A1* gene encodes antiquitin and is associated with Pyridoxine dependent epilepsy (antiquitin deficiency)[183]. Antiqutin is a member of subfamily 7 in the aldehyde dehydrogenase gene family and is thought to play a role in detoxification of aldehydes. Pyridoxine dependent epilepsy is autosomal recessive and reported variants tend to be compound heterozygous or homozygous. Up to 90% of cases present in the first month of life with clonic seizures and myoclonic jerks, often with irritability and hypotonia; although the clinical spectrum can extend to a later onset and a multisystem disorder[183,184]. Pyridoxine treatment causes cessation of seizures in the majority.

Subject S2294/00554 has seizure onset at 3 years 2 months with multiple seizure types; currently at the age of 13 years she has been seizure free for 7 years without Pyridoxine treatment. Her c.G1072A,p.E358K heterozygous variant had a MAF of 1.98 x 10^{-5} with gnomAD. She does not carry another *ALDH7A1* variant to enable a biallelic (compound heterozygous) presentation. Her phenotype and genotype are not consistent with Pyrdoxine dependent epilepsy.

HUWE1

HUWE1 encodes an E3 ubiquitin ligase, which is involved in neuronal development[185]. *HUWE1* is an X linked ID gene and was identified as it lies in the minimal overlap region of different

patients with Xp11.22 microduplication[186]. A dose sensitive relationship is inferred as cases demonstrate up regulation of *HUWE1* on expression array and quantitative PCR analysis[186]. The principal phenotypes are affected males with variable degree of ID and delayed speech acquisition. Dysmorphic features with abnormal growth parameters are sometimes present. Epilepsy is uncommon and seizure types are generalised or febrile. Carrier females may have evidence of ID[187]. The majority of families carry a microduplication, but four families with three segregation missense variants have been reported[186,188].

Subject S2341/00505 is a six year old male with seizure onset at 2 years. He has frequent and multiple seizure types but no associated ID. There is no family history of ID although a second-degree relative has seizures. He carries a novel splice site variant. His normal cognitive profile would not be in keeping with a *HUWE1* phenotype and the effect of the splice variant on dosage cannot be determined without further expression studies.

KIAA2022

KIAA2022 is an X-linked ID gene which was first described in hemizygous males[85]. Causative variants typically result in reduced *KIAA2022* expression and are truncating, frameshift or in one case a pericentric inversion causing disruption and in another case a gene duplication[85]. Symptomatic female carriers were then identified. Lange *et al.* presented 14 females with *de novo* heterozygous variants. ID was present in all and epilepsy in 12/14, including one patient with MAE (WTSI_MAE_1144306)[23].

Subject WTSI_MAE_1150147 is an 11-year-old female with seizure onset at 3 years 8 months of atonic seizures, which was her only seizure type. She has associated ID. There is no relevant family history. Whist her phenotype would not be inconsistent with *KIAA2022*, her variant is not frameshift or nonsense, is present in gnomAD, is inherited from her mother and *in silico* predictors are not supportive.

PLCB1

PLCB1 encodes the phosphoinositide-specific enzyme, phospholipase C β1. This enzymes generates the intracellular second messengers diacylglycerol and inositol-1,4,5-triphsophate from phosphatidylinositol-4,5-biphosphate. Biallelic loss of function variants in *PLCB1* has been described in four cases with early infantile EE. Two cases had a homozygous deletion affecting exons 1 to 3[189,190], the third has a compound heterozygous mutation in the same region[191], and the most recent published case has a homozygous deletion involving exons 7 to

8[192]. All cases developed seizures within the first year of life and have severe neurodevelopmental delay.

Subject WTSI_MAE_1150147 is an 11-year-old female with seizure onset at 3 years 8 months of atonic seizures, which was her only seizure type. She has associated ID. There is no relevant family history. She has a novel heterozygous splice site variant in exon 28 inherited from her mother. Exploration for other *PLCB1* variants showed that she had two other synonymous heterozygous variants, and a homozygous variant; MAF of these variants ranged from 0.05 to 0.9. Neither her phenotype or genotype is consistent with *PLCB1* pathogenic variants; hence this variant was classified as likely benign.

POLG

POLG encodes DNA polymerase gamma and is associated with five MIM mitochondrial disorders, three of which are associated with refractory seizures[193]. The seizure phenotypes are often occipital lobe seizures and status epilepticus. *POLG* related diseases have different degrees of severity and timing of onset, often with overlapping range of symptoms with multiple organ system involvement[193]. Symptoms include lactic acidosis, seizures, ataxia, peripheral neuropathy, developmental delay, myopathy, chronic progressive external ophthalmoplegia, and hepatopathy. A distinctive feature is the increasing involvement of neurological signs and symptoms during illness such as viral infections and physiological stressors[193]. Pathogenic variants are usually substitutions and homozygous or compound heterozygous and have been curated in an online database (http://tools/niehs.nih.gov/polg/). In a systematic review, 84% of cases with *POLG* related disease and epilepsy had one of three variants p.Ala467Thr, p.Trp748Ser and p.Gly848Ser[194].

Subject WTSI_MAE_1140147 has no other neurological features apart from epilepsy and ID. She carries a heterozygous p.1081_1083 deletion inherited from her mother. This deletion in not present in the *POLG* database and her only other *POLG* variant is a common homozygous variant. Neither her phenotype nor genotype is consistent with a *POLG* associated disease.

TBC1D24

TBC1D24 was first mapped in a consanguineous Arab Israeli family with focal epilepsy and ID using multipoint linkage analysis[84]. Since then, several other reports include a recent review on 48 cases has been published; this included a single case with MAE also recruited to this cohort. All reported *TBC1D24* variants are bialleleic and the related epilepsy syndromes show marked phenotypic pleiotropy[22].

Subject S1381 has a novel heterozygous variant with supportive *in silico* predictors, however she does not carry any other *TBC1D24* variants and hence does not fulfil the required bi allelic pattern.

7.6.2 Variants of uncertain significance

16 variants in 13 genes were identified (Table 7.8).

Table 7.8. Summary of variants of uncertain significance.

Sample ID	Gene	c.DNA:protein change	SIFT, PP2hvar, CADD or splice predictors	gnomAD MAF	Inheritance
S2298/ 00570	*ARHGEF9*	c.G1013A,p.R236Q	0.01 (P), 0.999 (P), 31	Novel	*De novo*
S1389/ 768H	*CACNA1A*	c.C6841T:p.R2281W	0.01 (P), 0.985 (P), 25.4	8.98×10^{-5}	NK
S1392/ 009Z	*CACNA1H*	c.C3833G:p.S1278C	0.03 (P), 0.765 (LP), 24.7	Novel	NK
S2342/ 00546*	*CHD2*	c.G3614A:p.R1295K	0.84 (T), 0.006 (T), 9.8	1.22×10^{-5}	Mother
S1395/ 295D	*CPA6*	G268T:p.G90C	0.05 (P), 0.54 (LP), 23.1	Novel	NK
S2332/ 00561	*FASN*	c.G5095C:p.V1699L	0.01 (P), 0.991 (P), 27.6	5.26×10^{-5}	NK
W_M_ 1150147	*GRIN2A*	c.G3827A:p.R1276H	-, 0.001 (T), 13.68	7.93×10^{-6}	Father
S1383/ 768H	*MTOR*	c.G5464A:p.G1822R	0.52 (T), 0.556 (LP), 23.6	1.24×10^{-5}	Mother
W_M_114 6134	*MTOR*	c.T1931C:p.V644A	0.39 (T), 0.481 (LP), 21.6	Novel	Father
S2340/ 00504	*PIK3R2*	c.A526G:p.S176G	0.58 (T), 0.001 (T) 0.85	Novel	NK
S1376/ 027A*	*SLC6A1*	c.953+6A>G	Ada 0.0001 (not sig.), RF 0.078 (not sig.)	Novel	Father
S1387/ 565D	*SPTAN1*	c.G4276C:p.E1426Q	0.37 (T), 0.958 (P), 17.8	Novel	NK
W_M_115 8865	*STX1B*	c.T662C:p.L221P	0 (P), 1.0 (P), 28.8	Novel	Mother
S2340/ 00504	*STX1B*	c.T845C:p.I282T 16:31004164A.G	0.01 (P), 0.104 (T), 23.1	7.97×10^{-6}	Mother
W_M_115 0147	*SYNGAP1*	c.A1925C:p.K642T	0 (P), 0.276 (T), 23.5	Novel	Mother
S2389 / 00529*	*SYNGAP1*	c.2177-2180del: p.726-727del	-	-	NK

W_M WTSI_MAE, T tolerated, LP likely pathogenic, P pathogenic, NK not known *case described in likely pathogenic section 7.6.3

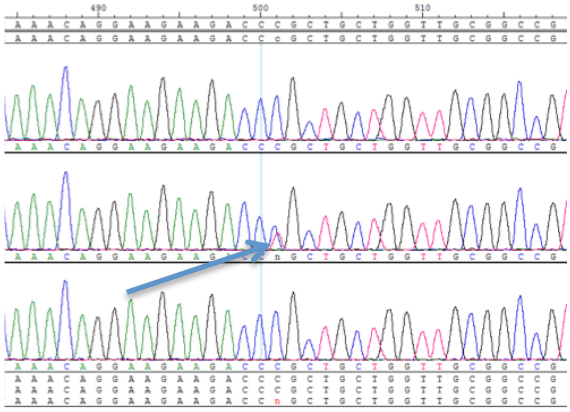
ARHGEF9

ARHGEF9 (Rho Guanine Nucleotide Exchange Factor 9) is an X linked gene that encodes Collybistin, which plays an important role in the localisation of gerphyrin, a scaffolding protein essential for postsynaptic clustering of GABA and Glycine receptors[195]. Harvey *et al*. published

the first case of a male with ID, epilepsy and hyperkeplexia and a p.G55A variant[196]. Since then 12 patients with *ARHGEF9* variants have been reported[197]. The clinical phenotype appears to be principally ID as well as seizures and psychiatric comorbidity in the majority, and variable dysmorphism. Only three female cases have been reported, two with evidence of X chromosome rearrangements. Marco *et al.* reported a female with ID and sensory hyperarousal with a *de novo* paracentric inversion of one X chromosome disrupting a single gene *ARHGEF9* and skewed inactivation of the normal X chromosome[198]. Kalscheuer *et al*. reported a female with a disturbed sleep wake cycle, seizures, anxiety and aggressive behaviour. She had a balanced chromosomal translocation with breakpoint on Xq11 disrupting *ARHGEF9*[199]. More recently an 8-year-old female with ASD, ID and speech delay was reported to carry a *de novo* 82kb deletion of chromosome Xq11.1-11.2 involving *ARHGEF9*[200], supporting the pathogenic role of *ARHGEF9* haploinsufficiency in females even with normal X chromosome arrangements.

Subject 00570 is a 9-year-old female with normal development prior to seizure onset at 12 months of GTCS, myoclonic atonic or atonic seizures, absence and focal seizures. She had ID and a diagnosis of ADHD. There was no family history of epilepsy or psychiatric disorders. The presence of seizures, ID and psychiatric comorbidity would be consistent with previously reported phenotypes. She carries a novel *de novo* missense variant (see Figure 7.12 for Sanger sequencing chromatogram perform by Universitätsklinikum Schleswig-Holstein Kiel) predicted pathogenic by *in silico* predictors. However her genotype association with *ARHGEF9* would be novel as no missense mutations in females have been reported, and the variants effect on haploinsufficiency is unknown. Based on the current evidence available, this variant was classified as a VUS.

Figure 7.12. Chromatogram demonstrating *de novo ARHGEF9* variant in subject 00570



Top sequence father, middle sequence child and bottom sequence mother.

CACNA1A

The *CACNA1A* gene encodes the alpha 1 subunit of the Cav2.1 P/Q type voltage gated calcium channel. *CACNA1A* variants are associated with three overlapping autosomal dominant conditions; episodic ataxia type 2 (OMIM 108500), spinocerebellar ataxia type 6 (OMIM 183086) and familial hemiplegic migraine (OMIM 141500). The relationship with epilepsy phenotypes is more complicating. Association studies linking an exon 8 polymorphism of *CACNA1A* with idiopathic generalized epilepsy were not replicated[201,202]. Still, seizures, particularly absence seizures are commonly seen in the three known associated conditions and in knockout mice[203]. *CACNA1A* loss of function mutations have also been found to segregate in four pedigrees with a spectrum of ID, autism, EE and cerebellar symptoms[204]. Additionally *de novo CACNA1A* variants have been identified in two cases of sporadic EE[205].

Subject S1389 is a nine year old Italian male with seizure onset at 9 months of atonic, myoclonic atonic and myoclonic seizures. He has been in seizure remission since 6 years but has moderate ID and autism. His mother and paternal uncle have epilepsy and ID. He has dysmorphic features: prominent forehead, small eyes, large mouth, high arched palate, syndactyly of first to third toes of both feet. EEG demonstrated GSW. His variant is supported by *in silico* prediction but is now reported in gnomAD. Given the family history, segregation studies will be most instructive in determining the role of this variant and Italian collaborators who have the parental DNA are planning to carry this out. Based on the current level of information, this variant was classified as a VUS.

CPA6

Carboxypeptidase A6 is an extracellular matrix metallocarboxypeptidase that removes hydrophobic C-terminal amino acids from peptides and proteins. The *CPA6* gene is highly expressed during brain development[206]. Homozygous variants in *CPA6* were first identified to be associated with familial febrile seizures and temporal lobe epilepsy in four siblings from healthy consanguineous Moroccan parents[207]. Subsequently, heterozygous missense variants were identified in a handful of temporal lobe epilepsy patients, although inheritance studies were not performed[208]. Allen *et al*. presented *CPA6* as a candidate gene in a 16 year old female with seizure onset at 5 months evolving into a LGS phenotype. She had two maternally inherited *CPA6* variants c.6919C>G and c.799G>A with a MAF <0.05[209].

Subject S1395 is a 15-year-old Italian female who presented with GTCS, myoclonic and clonic seizures when she was 3 years old. EEG showed left posterior sharp wave and GSW. She is currently in seizure remission and has no associated neurodevelopmental comorbidity. She has

a novel *CPA6* variant with conflicting support from *in-silco* predictors. Apart from its role as a private mutation, the evidence surrounding *CPA6* as causative epilepsy gene is still lacking.

FASN

Fatty acid synthase catalyzes the conversion of acetyl CoA and malonyl-CoA in the presence of NADPH (nicotinamide adenine dinucleotide phosphate), into long chain saturated fatty acids. Fatty acid synthase is essential in embryonic development and *FASN* deletion in mice results in impaired neurogenesis[210]. *FASN* was a statistically enriched gene for *de novo* variants in two individuals in a cohort of 365 US and UK trios with infantile spasms and LGS[155].

Subject 00561 is a 6-year-old female with onset of multiple seizure types at 2 years. She is currently in seizure remission and has no neurodevelopmental comorbidity. Her father has epilepsy. Her variant is present in gnomAD (MAF 5.26×10^{-5}) but has supporting *in-silico* scores. There is insufficient biological and functional work to extend the role of *FASN* in epilepsy beyond the role of candidate at present. Inheritance studies were not available.

GRIN2A

GRIN2A encodes the N methyl D aspartate (NMDA) glutamate receptor α2 subunit which is involved in the long term potentiation of synaptic transmission. *GRIN2A* variants are associated with idiopathic focal epilepsy. The principal phenotypes consist of an element of speech and language impairment; namely the epilepsy aphasia syndromes (Landau-Kleffner syndrome, continuous spike and wave during slow wave sleep), rolandic epilepsy and atypical benign partial epilepsy[211-213]. *GRIN2A* variants were not identified in 475 cases with EE including 85 cases with MAE[211]. However three cases of EE have been published with *de novo* *GRIN2A* variants[214-216], and an adjunctive trial of memantine was able to reduce seizure burden in one of these cases with a gain of function mutation[216]. Both *de novo* and inherited variants sometimes with incomplete penetrance have been identified in *GRIN2A*. All types of mutation (missense, nonsense, splice site, deletions) have been reported and appear to cluster in the S1 and S1 peptide segments of the ligand binding domain although can occur throughout the protein.

Subject WTSI_MAE_1150147 has severe MAE with developmental stagnation following seizure onset. Her variant is inherited from an unaffected father. It is located in the C-terminal domain and is in the identical location but with an alternate allele to a reported possibly pathogenic variant p.A1276G in a patient with continuous spike wave in slow wave sleep. The authors recognised that the variant was present in control populations and inheritance was not

established in their case[212]. Given that *GRIN2A i*s usually associated with a focal epilepsy phenotype, along with no other strongly supporting features in genotype (present in gnomAD, located in C terminal domain, inherited from unaffected parent, *in silico* predictors not supportive); this variant was classified as a VUS.

MTOR

The *MTOR* gene is part of the mTOR (mammalian target of rapamycin) pathway, which is an essential pathway that regulates cell growth in response to a wide range of stimuli. Somatic *MTOR* mutations were first identified to cause focal cortical dysplasia and intractable epilepsy from exome sequencing of paired blood-brain DNA samples[217]. Low frequency somatic variants reported had allelic read frequencies typically less than 10% (range 1.26% to 12.6%)[217]. Recently, data sharing amongst European and US investigators led to the identification of 12 patients with germline *MTOR* variants. All germline variants identified were either *de novo* or demonstrated segregation with parent to child transmission. 10/12 of these cases were classified with focal epilepsy with MRI brain abnormalities findings in nine. One case with a *de novo c*.C478T,p.M1595I variant had a generalised epilepsy at onset of 12 months with myoclonic and GTCS. He had a history of developmental delay prior to seizure onset and has moderate ID and autism. EEG demonstrated a slow background with generalised 1 -2 Hz slow spike wave[218].

Subject WTSI_MAE_1146134 is a 6-year-old male with seizure onset at 2 years with myoclonic, myoclonic-atonic and GTCS. EEG demonstrated generalised epileptic activity. He has no neurodevelopmental comorbidity and has no relevant family history. He carries an inherited missense variant, which is unlikely to be somatic as alternate read counts are 44.6% (54 out of 121). His phenotype is not typical for *MTOR* germline variants and does not demonstrate segregation; hence it is classified as a VUS.

Subject S1383/768H is an Italian male with severe MAE, moderate ID and dysmorphic features with seizure onset at 9 months. EEG shows GSW. There are no neuroimaging abnormalities of note. Significantly his mother also has epilepsy and ID. Subject S1383 carries a missense variant inherited from his mother, it is identified in gnomAD and *in silico* predictors are partially supportive. Alternate read frequency is 53.2% (65 out of 122), not suggestive of a somatic mutation. As a germline inherited mutation, the phenotypic segregation is supportive, however this variant is now seen in population databases MAF 1.24×10^{-5}, hence classified as a VUS.

PIK3R2

PIK3R2 (phosphoinositide-3-kinase regulatory subunit) encodes the p85b regulatory subunit of the phosphatidylinositol 3-kinase-AKT pathway[219]. Genes within the phosphatidylinositol 3-kinase-AKT pathway are known to cause a wide range of developmental brain and body disorders. *PIK3R2* is associated principally with megalaencephaly and bilateral perisylvian polymicroqyria, both conditions that lead to epilepsy. About a dozen patients with *PIK3R2* variants have been reported. Variants are *de novo* and can be heterozygous or mosaic, with variable alternate allele read levels ranging from 2% to 37%[220,221]. Three recurrent variants p.G3733R, p.D557H and p.L401P account for the majority of reported cases and cluster in the SH2 domain. A single case with a *de novo* p.D557H variant in the PI3K catalytic subunit α has been reported[222].

Subject S2340/00504 is a 17-year-old female with severe MAE and ID. She has no abnormalities on brain neuroimaging. She has a novel missense variant in the Rho GAP domain, an area not previously associated with pathogenic variants. The alternate read ratio for the variant is 46.8% (22 out of 47). This variant location along with lack of a co-existing brain malformation and lack of inheritance studies leads to classification of this variant as a VUS.

SPTAN1

SPTAN1 which encodes α II spectrin was first identified as mutated in two patients with West syndrome by Japanese investigators[223]. To date, a total of seven epileptic patients with *SPTAN1* variants have been identified and this cohort offers a distinct phenotype and genotype. The clinical phenotype includes EE with hypsarrhythmia (West syndrome), no visual attention, acquired microcephaly, spastic quadriplegia and profound ID. A single case had a milder phenotype with well-controlled generalised epilepsy but still profound ID[224]. MRI brain abnormalities demonstrating brainstem and cerebellar atrophy and cerebral hypomyelination are present and offer additional specific clues to this gene diagnosis. All pathogenic variants reported are *de novo* in-frame in the last two of 20 spectrin repeats in the C-terminal region. The last four spectrin repeats in the C-terminal region are required for α/β spectrin heterodrimer associations and these mutations alter heterodimer formation between the spectrins[225]. A *de novo* missense variant p.R566P in spectrin repeat 5 of unclear significance has been reported in a patient with non syndrome ID and no epilepsy[224].

Subject S1396/565D is a 10-year-old Italian boy with seizure onset at 11 months. His seizure types include absence, atonic, tonic and myoclonic. He has no evidence of cognitive decline.

Unfortunately EEG and neuroimaging findings are not available for this case. He carries a novel missense variant located in spectrin repeat 5. Inheritance studies were not available. His phenotype and genotype are not consistent with typical *SPTAN1* pathogenic variants and therefore this variant was classified as a VUS.

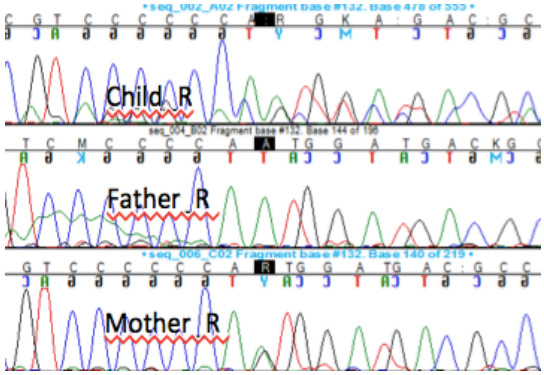
STX1B

STX1B encodes syntatxin 1b, a component of the SNARE complex, that helps tether synaptic vesicles at the presynaptic membrane and mediates neurotransmitter release[226]. *STX1B* as a causative gene was first identified in two large German pedigrees with familial febrile seizures. Linkage analysis isolated a single locus on chromosome 16p11.2 with a LOD score of 4.27 leading to the identification of *STX1B* following exome sequencing more than a decade later[20]. Subsequently seven cases with *STXIB* variants have been reported[20,82]. These variants tend to cluster in the SNARE motif and were missense, truncation or deletions. Of note co-segregation was not complete and the pathogenic variant was present in one unaffected individual in the linkage family. Three of these seven reported cases have a MAE phenotype, including a case (WTSI_MAE_1199460) from this cohort with a *de novo* cG676C,p.G226R variant, the other two have a *de novo* 0.8Mb and 1.2 Mb microdeletion with 160kb overlap encompassing all or part of *STX1B*, suggesting a role for haploinsufficiency in pathogenesis.

Subject WTSI_MAE_1158865 from Antwerp carries a novel missense variant located in the SNARE motif, inherited from mother. *In silico* predictors are supportive. The patient has severe epilepsy with overlapping features of MAE and LGS with onset at 3 years. There is no history of febrile convulsions. She has mild ID. Further investigation of family history revealed a history of seizures in the maternal grandfather who is deceased; there is no history of seizures in the mother (personal communication Dr Sarah Weckhuysen). Given the severity of subject WTSI_MAE_1158865's phenotype, it is difficult to define the exact role of this inherited variant in spite of the other supporting information.

Subject S2340/00504 is 17-year-old female with seizure onset at 19 months. She has severe MAE with multiple seizure types and moderate to severe ID. There is no history of febrile seizures and no family history of epilepsy. Her maternally inherited *STX1B* p.I282T variant is located close but not inside the SNARE motif (ends position 252). (see Figure 7.13). As there is no evidence of phenotypic segregation, this variant was classified as a VUS.

Figure 7.13. Chromatogram demonstrating maternally inherited *STXIB* variant in subject 00504.



R denotes sequences obtained by reverse primer.

7.6.3 Likely pathogenic variants

Eight subjects had novel variants, which were classified as likely pathogenic within seven genes *CHD2, KCNH5*, *KCNB1, KIAA2022, MECP2, SLC6A1* and *SYNGAP1* (see Table 7.9). Of these six were *de novo* variants and in the other two (*KIAA2022, SYNGAP1*), inheritance was not known. The occurrence of a *de novo* variant in a known epilepsy gene is statistically very unlikely to happen by chance but cannot be taken as an implicit assumption of pathogenicity. Therefore phenotypic correlations and pathogenic variant characteristics with previously reported cases are vigorously explored.

Some of the work presented here were done elsewhere. The *MECP2* variant in subject 291J and *SYNGAP1* variant in subject 00512 was identified here, but Sanger validation was performed by Italian and German collaborators as they had parental DNA.

Table 7.9. Summary of likely pathogenic epilepsy variants.

Sample ID	Gene	c.DNA:protein change	SIFT, PP2hvar, CADD or splice predictors	gnomAD MAF	Inheritance
S2331/ 00559	*CHD2*	c.3734+7A>G	-	Novel	*De novo*
S2393/ 00533	*KCNB1*	c.C916T:p.R306C	0 (P), 1.0 (P). 29.7	Novel	*De novo*
S2391/ 00525	*KCNH5*	c.1569+7G>T	Ada 0.0001 (not sig.), RF 0.004 (not sig.)	Novel	*De novo*
S1388/ 138J	*KIAA2022*	c.1261_1270del: p.L421fs	-	Novel	NK
S1394/ 291J	*MECP2*	c.C673A:pP225T	0.05 (P) 0.998 (P), 24.4	Novel	*De novo*
S2537/ 00595	*SLC6A1*	c.C1155G:p.F385L	0.03 (P), 0.012 (T), 32	Novel	*De novo*
S2389/ 00512	*SYNGAP1*	c.2176_2179del: pR726fs	-	Novel	*De novo*
S2287/ 00514	*SYNGAP1*	c.2562_2578del: p.854_860del	-	Novel	NK

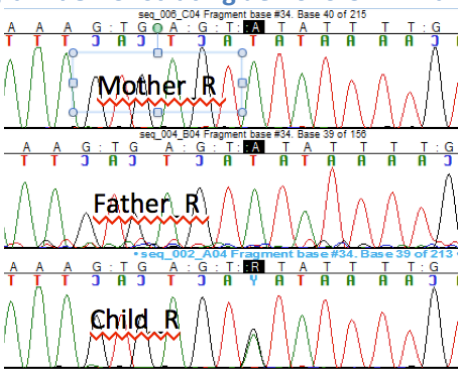
P pathogenic, T tolerated, NK not known

CHD2

CHD2 (Chromodomain helicase DNA-binding protein) take the position as the first MAE gene identified in the era of NGS. *CHD2* belongs to the SNF2 related superfamily of ATPase and plays a pivotal role in modulating chromatin structure. It was initially identified as a candidate gene due to its location in a *de novo* 15q26.1 rare CNV in epilepsy[227]. Subsequently, *De novo CHD2* variants have been identified in individuals with MAE, LGS and EE[14,15,77]. The phenotype is consistent with seizure onset usually between 2 to 3 years of myoclonic seizures, with myoclonic absence or myoclonic atonic seizures and induced seizures (fever or photosensitivity). ID is invariably seen although the severity may vary. All published pathogenic *CHD2* variants are *de novo* and have been located throughout the protein; mutations can be frameshift, missense, nonsense or splice site.

Subject S2331/00559 is an eight-year-old female; she had normal development prior to seizure onset at 2 years 6 months with GTCS, myoclonic atonic and myoclonic absences seizures. EEG demonstrated polyspike and wave activity. She has ID and ADHD. There is no family history. Her novel splice site variant at c.3734+7A>G was validated and found to be *de novo* (Figure 7.14). Her phenotype and genotype would be consistent with *CHD2* associated epilepsy.

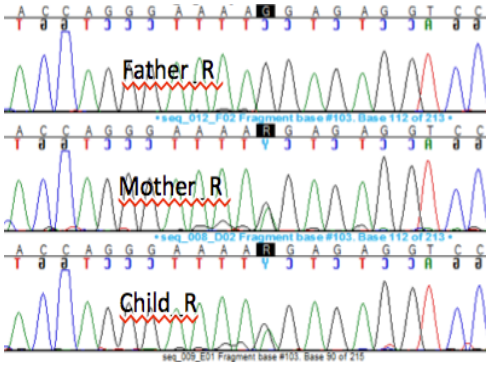
Figure 7.14. Chromatogram demonstrating *de novo CHD2* variant in subject 00559



R denotes sequences obtained by reverse primer.

Subject S2342/00546 is a 12 year old with previously normal development and seizure onset at 4 years 7 months with GTCS, myoclonic atonic, myoclonic and absence seizures. She has no associated neurodevelopmental co-morbidity and no family history. She carries a maternally inherited *CHD2* variant (Figure 7.15), gnomAD MAF 1.22 x 10^{-5}. This variant was classified as a VUS.

Figure 7.15. Chromatogram demonstrating maternal inheritance of *CHD2* variant in subject 00546.

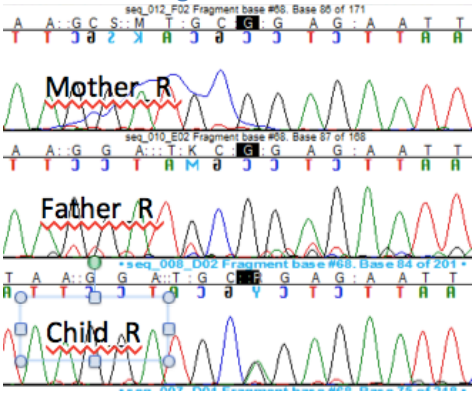


R denotes sequences obtained by reverse primer.

KCNB1

KCNB1 (voltage gated potassium channel subfamily B member 1) encodes the Kv2.1 voltage gated potassium channel that mediates transmembrane potassium transport in excitable membranes, primarily in the brain, but also in the pancreas and cardiovascular system. Allen *et al.* first identified a single *de novo* case with a p.T374I variant in a 5 year old female with EE and cerebral palsy[101]. Since then five further patients with EE and infantile epilepsy have been reported with *de novo KCNB1* variants. Functional studies have demonstrated a deleterious effects on the Kv2.1 mutation on loss of ion selectivity and gain of a depolarizing inward cation conductance[228]. One of the published patients had the same variant as subject S2393/00533 described here, which was established as *de novo*[228,229] (see Figure 7.16).

Figure 7.16. Chromatogram demonstrating *de novo KCNB1* variant in subject 00533.



R denotes sequences obtained by reverse primer.

The clinical features of these two cases with a *de novo* p.R306 variant is summarised in Table 7.10. They both have an early onset drug resistant epilepsy associated with severe ID. There were no similarities of specific seizure types or EEG features in these two cases but atonic seizures and multifocal spikes on EEG seen in subject S2393/00533 have been reported in other patients with pathogenic *KCNB1* variants. Additionally it can be difficult to ascertain prior developmental delay in subject S2393/00533 whose seizure onset was at 6 months. This recurrent *de novo* variant was considered pathogenic.

Table 7.10. Clinical features of subject 00533 and other reported case with *KCNB1* p.R306C variant.

	Patient 2[229]	**Subject 00533**
Variant	c.916C>T, p.R306C	c.916C>T, p.R306C
Age and Sex	7 year old Male	6 year old Female
Prior development	Developmental delay	Normal
Age at seizure onset	1 year	6 months
Seizure types	Spasms, tonic clonic, myoclonic and focal seizures	Myoclonic atonic, myoclonic, atonic,
EEG findings	Generalised spike and polyspike wave discharges	Frequent and brief spikes with no particular focus
Response to therapy	Refractory	Refractory
Intellectual disability	Severe	Severe
Other clinical features	Tantrum burst, macrocephaly	-

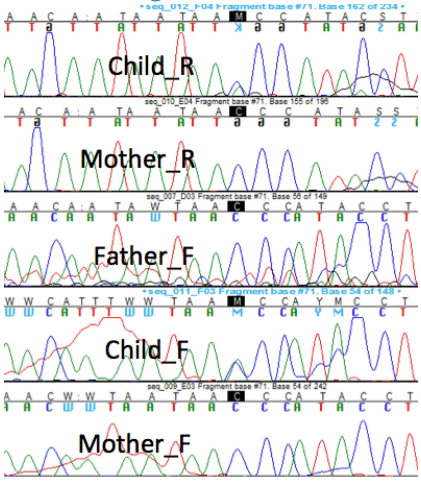
Paternity for this trio was confirmed through Sanger sequencing of homozygous *LOC4000863* variant which was present in the child (S2393/00533), and heterozygous in both parents.

KCNH5

The *KCNH5* (voltage gated potassium channel Kv10.2) gene encodes the Kv10.2 channel, part of the ether-a-go-go family within the voltage gated potassium channel superfamily. Kv10.2 channels are natively expressed in the nervous system but its function is not well understood. Only one case with a *de novo* c.G980A,p.R327H in *KCNH5* has been reported, the authors considered this variant a good candidate[230]. The patient was a 13-year-old male with seizure onset at 6 months. Seizure types were GTCS or hemiclonic with occasional brief facial clonic seizures. EEG demonstrated frequent multifocal spikes. Neurodevelopmental difficulties of autism, developmental delay and ID were present. Subsequently functional work on this same variant with voltage clamp analysis demonstrated hyperpolarizing shift of voltage dependence of activation and thus the role of this residue in destabilising the channel-closed state[231].

Subject S2391/00525 is a 5-year-old female. There was early developmental delay followed by seizure onset at 3 years with GTCS. Subsequently she developed multiple seizure types consisting of atonic or myoclonic atonic, myoclonic, absence and tonic seizures. She continues to have drug resistant epilepsy with daily to weekly seizures. She has moderate to severe ID. EEG shows frequent generalised and focal interictal epileptiform activity. She also suffers from recurrent chest infections, constipation and sleep difficulties. A *de novo* splice site variant c.1569+7G>T in *KCNH5* was identified (see Figure 7.17).

Figure 7.17. Chromatogram demonstrating *de novo KCNH5* variant in subject 00525



R denotes sequences obtained by reverse primer, F denotes sequences obtained by forward primer

These two cases have similarities in the form of complicated epilepsy, associated ID and focal features on EEG, which in Subject S2391/00525 is an unusual finding in MAE. There are no available further supportive features based on the variant type, location or with *in silico* predictors. Blood RNA samples are currently being collected with plans to perform mRNA

expression studies to determine the effect on the transcript. Altogether with the evidence that is available, this variant is classified as likely pathogenic.

KIAA2022

See section 7.6.1 for background on this gene.

Subject S1388/138J is a previously normal developing 10-year-old Italian female who had her first seizure at 21 months with myoclonic atonic seizures. EEG demonstrated GSW. She has associated mild to moderate ID. She has a frameshift variant in exon 3, a genomic 'hotspot' where most of reported pathogenic variants are located. Her phenotype consisting of ID and seizures is also supportive. Italian collaborators are now carrying out inheritance studies. This variant has been classified as likely pathogenic in view of the supportive phenotypic and variant characteristics.

MECP2

The *MECP2* (methyl-CpG-binding protein 2) gene is an X-linked gene associated with Rett syndrome (OMIM 3127250). Classic Rett syndrome is easily identifiable to clinicians and occurs mainly in girls. It is a severe neurodevelopmental disorder characterised by arrested development in the first 18 months of age with regression of skills and speech, stereotypic movements, microcephaly, ID and seizures. Seizures were present in 94% in a series of 54 cases, with focal seizures with secondary features most common[232]. Whilst no unique or diagnostic EEG features exist, certain EEG findings are more common. There is a loss of occipital dominant rhythm and a slow background with theta and delta activity, and both generalised and focal electrographic seizures are seen[233].

Atypical forms of Rett syndrome with an either milder or more severe phenotype exist and revised diagnostic criteria differentiating classic and atypical forms have been proposed[232]. Other phenotypes associated with *MECP2* variants include non-specific ID, autism and neonatal encephaloathy[233]. Pathogenic variants are *de novo* in 99% of Rett syndrome and the eight most common variants account for 50% of cases. Missense, nonsense, splice site and small indels are dispersed throughout the gene have been reported and are curated in various open access database such as RettBASE (http://www.mecp2.chw.edu.au).

Subject S1395/291J is a 15-year-old Italian female with a previous history of febrile convulsions. Her seizure onset was at 6 years with myoclonic atonic and tonic seizures. She has moderate to severe ID and cerebellar signs on clinical examination. She is currently in

seizure remission. EEG shows disorganised background activity with bilateral paroxysmal activity. Her novel *de novo* variant p.P225T is in the transcriptional repression domain in the same position but a different allele from p.P225L in a male patient with ID and autism[234]. Although her phenotype would not fall into the diagnostic criteria for classic or atypical Rett syndrome, she has cerebellar signs and disorganised background on EEG, which are consistent with a Rett syndrome phenotype and are unusual for MAE. This variant was classified as likely pathogenic. A possible clinical implication is the awareness of prolonged QT_c (QT corrected interval) syndrome in patients with Rett syndrome regardless of the severity of their phenotype, such that patients with Rett syndrome should avoid drugs that cause prolonged QT_c syndrome[235].

SLC6A1

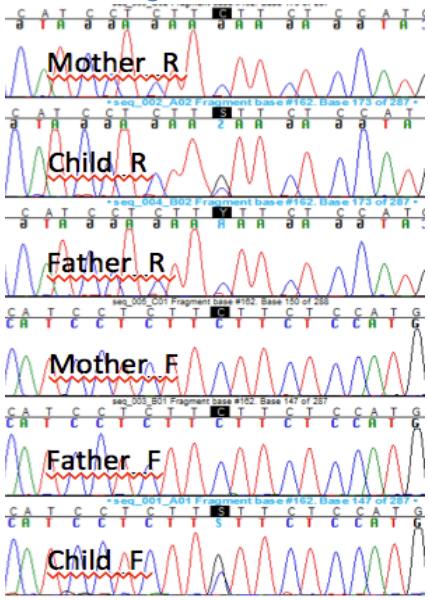
SLC6A1 encodes GAT-1, a voltage dependent gamma aminobutyric acid (GABA) transporter that is responsible for the re-uptake of GABA from the synapse. *SLC6A1* has been described as the Doose gene as it has been linked only to MAE patients and at 4% (6/160 cases)[17], it is the most enriched causative MAE gene. Carvill *et al.* published six pathogenic *de novo* heterozygous *SLC6A1* variants, two truncation and four missense[17]. The authors postulated a loss of function role in the nonsense variants and clustering of the four missense variants around the GABA binding pocket which in three dimensional space disrupts GABA transport from the extracellular space into the pre-synaptic terminal. Subsequently another *de novo* missense variant in p.C164T was identified[75]. The authors performed high resolution structural modelling to demonstrate disruption of a disulphide bond necessary for stabilization of the transmembrane helix. No functional or animal studies have been performed.

SLC6A1 associated MAE tends be severe, developmental delay is usually present prior to seizure onset, and subsequent ID of varying severity is present. Seizure types are atonic as well as myoclonic and absences. EEG shows generalised epileptic activity. Psychiatric comorbidity is present in the majority.

Subject 00595 is an 8 year old with prior development delay and had his first seizure at 13 months, which was a myoclonic seizure. Subsequently he developed myoclonic atonic seizures but has been in seizure remission for the past 2 years. He has ID and autism. There is no family history of epilepsy. His EEG demonstrated an unusual feature of eye closure sensitivity with GSW on the posterior third of the head. He was found to have a novel *de novo* p.F385L *SLC6A1* variant (see Figure 7.18). His phenotypic features of delayed development prior to seizure onset, ID and autism are consistent with a *SLC6A1* phenotype. His history of seizure offset has

also been reported in 3/7 of the other published cases. Additionally his unusual EEG feature of posterior eye closure sensitivity was seen in another *SLC6A1* patient (individual one, Carvill *et al.*[17]). This variant was classified as likely pathogenic.

Figure 7.18 Chromatogram demonstrating *de novo SLC6A1* variant in subject 00595.



R denotes sequences obtained by reverse primer and F denotes sequences obtained by forward primer.

Subject S1376/027A is an 18-year-old male with moderately severe MAE with onset at 2 years 5 months. He has mild ID and hyperactive behaviour. He carries a novel splice site variant within the GABA binding pocket. Given the unique association between *SLC6A1* and MAE, these results were communicated and Italian collaborators performed validation and inheritance studies. The variant was validated and inherited from an unaffected father (personal communication Dr. Davide Mei, Professor Renzo Guerrini). Given the lack of segregation, it was classified as a VUS.

SYNGAP1

De novo SYNGAP1 variants were first identified in 1% of 500 individuals with EE[227]. The phenotype is characterised by early neurodevelopmental delay typically preceding onset of generalised seizures consisting of absences and myoclonic jerks. Seizure triggers appear common and were present in 7/16 in a recent cohort (photosensitivity in five, fixation off sensitivity in two and chewing in one)[16]. Hypotonia and unstable gait are frequently associated neurological features. Pathogenic variants are *de novo* and mostly nonsense or frameshift; leading to a premature termination codon in the protein sequence. A single *de novo* missense variant was included in a cohort of 16 cases. Variants are located throughout the protein but

patients with pathogenic variants in exons 8 to 15 tend to have a more pharmacoresistant epilepsy[16].

Four subjects with novel *SYNGAP1* variants were identified. Subject 00514/S2287 is a 5-year-old male with a history of speech delay who developed myoclonic atonic, atonic and absence seizures at 17 months. His epilepsy evolved into a refractory epilepsy with severe ID and speech delay. He has clinical signs of hypotonia and ataxia. EEG shows GSW with diffuse background slowing. He carries a frameshift deletion in exon 15. Both his phenotype and genotype is consistent with *SYNGAP1* encephalopathy, therefore it is classified as likely pathogenic. Unfortunately I was unable to carry out Sanger sequencing validation studies for subject S2287/00514 due to insufficient DNA. DNA is currently being collected again to carry out validation and inheritance studies.

Subject 00512/S2389 is a 3 year 9 month old male with a background of global developmental delay who presented with myoclonic atonic and absence seizures at 2 years. He has severe ID, significant speech delay and autism. He walks with an ataxic gait. EEG shows GSW. He carries a frameshift deletion in exon 13. This variant was validated and identified as *de novo* by collaborators in Kiel, Germany (personal communication Dr Manuela Pendziwiat). Like subject S2287/00514, his phenotype and genotype is consistent with *SYNGAP1* encephalopathy, therefore it is classified as likely pathogenic.

Subject 00529/S2389 is a previously normally developing 6-year-old female with seizure onset at 5 years with multiple seizure types. She has no neurological signs and no ID. EEG shows polyspike and wave. Her genotype involving a frameshift deletion in exon 13 is supportive but the lack of ID renders her variant to be classified as a VUS. Re-assessment of her phenotype at a later date and inheritance studies may lead to variant classification change.

Subject WTSI_MAE_1150147 is a previously normally developing 11-year-old female with seizure onset at 3 years of myoclonic atonic seizures. She has reported ID. There is no family history of epilepsy. She carries a novel inherited missense variant from her mother. The absence of segregation and type of variant along with her phenotype would not be consistent with previously reported cases. This variant was classified as a VUS.

7.7 Gene exploration using aetiologically relevant gene sets

7.7.1 Neuropsychiatric gene set

Gene matching with Neuropsychiatric gene set

Twenty genes from the neuropsychiatric gene set had recurrent CADD>20 non-synonymous variants in the MAE cohort. There were no identical variant matches in the MAE cohort and the gene set variants. Additional filtering was carried out in order to select the best candidate genes for Sanger validation and inheritance studies. The additional filtering results were as follows; ten genes were de-prioritised due to RVIS >25th centile and negative Z score (*CNTN5*, *CUBN*, *MARK1*, *MKI67*, *MYOM3*, *SHANK2*, *TACC2*, *TET1*, *USH2A*, *WDFY4*), four genes were de-prioritised due to conflicting gene function and/or inadequate nervous system expression (*BAGALNT4*, *LRP4*, *MDM2*, *NCOR2*), three genes were de-prioritised due to variant or heterozygosity types (*AHDC1*, *ANK3*, *BRSK2*) and one gene was de-prioritised as the subject had inherited the variant from an unaffected parent (*NAV2*). Hence, two genes remaining were listed as possible candidates (Table 7.11).

Table 7.11. Candidate genes from gene matching with neuropsychiatric gene set.

Gene RVIS/ExAC Z	NP paper with variant and MAE subject ID	cDNA:protein change and inheritance if known	SIFT,PP2har,CADD / gnomAD MAF
ASH1L 2.22%/3.05	ID_DeLigt[175] S2392/00530 W_M_1202321	c. G2170T:p.A724S c.C4024T:p.R1342X c.A2578G:p.K860E inh M	 1 (T), -, 38 / Novel 0.01 (T), 0.987 (P), 38 / Novel
CHD4 2.82%/7.05	Epi4K[101] DDD[162] DDD[162] S1883/00526 W_M_1152372 W_M_1158865	c.G1000T:p.R334S c.C3360T:p.R1120Q c.C1379T:p.C460Y c.A2687G:p.H896R c.C3898T:p.R1300W inh F c.G4042A:pG1348S inh M	 0 (P), 0.993 (P). 24.6 / Novel 0 (P), 0.999 (P), 35 / Novel 0.15 (T), 0.946 (P), 25 / Novel

NP neuropsychiatric, W_M WTSI_MAE, inh inherited, M mother, F father, T tolerated, P pathogenic

Exploration of possible candidates from neuropsychiatric gene set

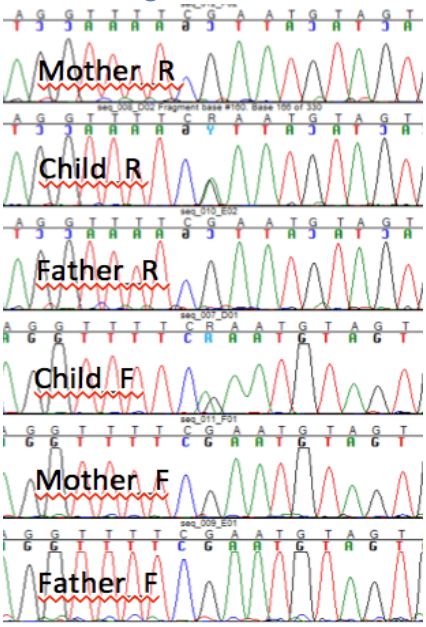
ASH1L

ASH1L (ASH1 Like Histone Lysine Methyltransferase) encodes a member of the trithorax group of transcriptional activators. Ash1L is widely expressed in multiple organs and enriched in the brain in mice. It is localised to nuclear speckles and also in cell-to-cell tight junctions. A role in epigenetic modification in brain functioning was implicated when Ash1L knockout mice completely abolished the activity dependent repression of neurexin 1α, a presynaptic adhesion

molecule required for synaptic formation[236]. More noteworthy is that gene ontology annotation for *ASH1L* includes chromatin binding, a pathway that has been evident in MAE with *CHD2* and *SMARCA2* genes.

Two cases from the MAE cohort here had *ASH1L* variants with CADD scores >20. Both variants are novel on gnomAD. WTSI_MAE_1202321 had a maternally inherited variant; there is a family history of epilepsy but only in the paternal uncle. Subject 00530/S2392 is a 7 year old male with seizure onset at 6 months. He has multiple seizure types: myoclonic atonic, myoclonic, GTCS, absence, tonic and they are refractory. EEG shows GSW and polyspike as well as focal activity maximal in the right temporal region. He has moderate to severe ID, ASD and ADHD. There is no family history of epilepsy. He has a p.R1342X *ASH1L* variant. Sanger sequencing and inheritance studies demonstrated that it was *de novo* (see Figure 7.19), strengthening the case for this gene to be a possible pathogenic variant. The introduction of *ASH1L* as an epilepsy candidate gene will be discussed in chapter 8, section 8.3.4.

Figure 7.19 Chromatogram demonstrating *de novo ASH1L* variant in subject 00530.



R denotes sequences obtained by reverse primer, F denotes sequences obtained by forward primer.

CHD4

CHD4 (chromodomain helicase DNA binding protein 4) is an ATP dependent chromatin remodeler involved in epigenetic regulation of gene transcription, DNA repair and cell cycle progression[237]. It is also a paralogue of *CHD2*[237], which has been associated with MAE. *De novo CHD4* variants have been identified in one case with EE[101] and two cases in the DDD study[162]. Individual phenotypes were not routinely available in the DDD study and were requested, both

of these cases did not have seizures (personal communication Dr. Katherine Lachlan). Most recently five *de novo* substitutions were identified in individuals with overlapping phenotypes of developmental delay, ID, hearing loss, macrocephaly, distinct facial dysmorphisms and hypogonadism[238]. On further enquiry with the authors, it was confirmed that none of these cases had seizures (personal communication Dr. Melita Irving).

Three MAE cases had CADD score >20 *CHD4* variants. All variants were novel in gnomAD. Subject WTSI_MAE_1152372 has a missense variant p.R1300W inherited from an unaffected father with no further family history of epilepsy. Subject WTSI_MAE_1158865 had a maternally inherited p.G1348S variant, family history of epilepsy was described as possible in mother and maternal uncle. Subject S1883/00526 is a 6 year old female from consanguineous unaffected Bangladeshi parents with a p.H896R variant (SIFT 0 (pathogenic), PP2hvar 0.993 (pathogenic), CADD 24.6). She has severe MAE and severe ID. Unfortunately we were not able to collect paternal DNA. Sanger sequencing was attempted on the proband and her mother but the region proved difficult to amplify with standard and optimised PCRs with four different sets of primers. A possible solution would be to perform a nested PCR for this region.

CHD4 is already proven to be associated with a neurodevelopmental phenotype but its role in seizure disorders and specifically MAE is still lacking, it remains a good candidate gene for further studies.

Enrichment analysis with neuropsychiatric genes

Overall, 193 (mean 2.44) novel CADD>20 neuropsychiatric gene set variants were identified in 79 cases compared with 10773 (mean 0.32) out of ExAC_NFE 33370 controls. Per gene, 96 genes in the neuropsychiatric gene set reached statistical significance set at $P<2 \times 10^{-5}$ for enrichment of novel CADD>20 variants in the MAE cohort. Table 7.12 lists the 10 most enriched genes in the MAE cohort compared to the ExAC_NFE cohort. Nine genes in this list overlapped with the de-prioritised genes identified from direct gene matching. The only gene not previously reviewed was *CDH23*. *CDH23* (Cadherin Related 23) is part of the cadherin superfamily and encodes calcium dependent cell adhesion glycoproteins. The encoded protein is involved in proper organisation of the sterocilia bundle of hair cells in the cochlea and the vestibule and *CDH23* has been linked to a spectrum of hearing loss diseases[239]. The ExAC Z score for *CDH23* is -0.24, RVIS 2.06%. Based on this information, *CDH23* was not considered a good candidate gene for MAE.

Table 7.12. Top 10 genes with smallest *P* values identified through enrichment analysis with neuropsychiatric gene set.

Gene	MAE_pos	MAE_neg	NFE_pos	NFE_neg	Chi-square	1 tailed *P* value
OBSCN	12	67	0	33370	5070.6798	$< 1 \times 10^{-5}$
CDH23	6	73	0	33370	2534.8850	$< 1 \times 10^{-5}$
AHDC1	4	75	0	33370	1689.8223	$< 1 \times 10^{-5}$
WDFY4	4	75	0	33370	1689.8223	$< 1 \times 10^{-5}$
BRSK2	3	76	0	33370	1267.3288	$< 1 \times 10^{-5}$
MYOM3	3	76	0	33370	1267.3288	$< 1 \times 10^{-5}$
BAGALNT4	2	77	0	33370	844.86064	$< 1 \times 10^{-5}$
CNTN5	2	77	0	33370	844.86064	$< 1 \times 10^{-5}$
MDM2	2	77	0	33370	844.86064	$< 1 \times 10^{-5}$
NCOR2	2	77	0	33370	844.86064	$< 1 \times 10^{-5}$

MAE = this exome sequencing cohort, pos = no. of times novel CADD>20 gene variants identified, neg = no. of times novel CADD<20 gene variants not identified, NFE non Finnish European. Level of significance $P<2 \times 10^{-5}$.

7.7.2 Ion channel gene set

Overall, 31 (mean 0.39) novel CADD>20 ion channel gene set variants were identified in 79 cases compared with 1103 (mean 0.033) out of ExAC_NFE 33370 controls. Direct gene matching identified 30 genes and within this group an overlap of 11 genes achieved $P<2 \times 10^{-4}$ following per gene chi square comparisons. All 30 genes were reviewed.

Additional filtering of genes was as follows: four genes (*CACNA1H, GRIN2A, KCNB1, SCN1A*), were de-prioritised as they are known epilepsy associated genes and were reviewed already. Nine genes (*BSND*, *CACNA1F*, *CHRNA5, GRM2, HCN3, HTR2A, HTR4, KCNA3, KCNS2*) were de-prioritized due to RVIS or ExAC score, five genes (*CACNA1S, GLRA1, RYR1, SCN10A, SCN11A*) were de-prioritised due to inadequate nervous system expression, one gene (*KCNJ12*) was removed as it was a calling error on IGV and seven genes (*GRID1, GRIK4, GRM3, CACNA2D3, CACNB2, CLCN1, SLC12A5*) were de-prioritised due to demonstrated inheritance of the variants with unaffected parents. One subject, WTSI_MAE_1147805 carried a maternally inherited *KCNAB1* variant from a mother with a possible history of seizures. *KCNAB1* has a Z score of 1.82 and borderline RVIS score of 24.46% and was not pursued further as a candidate. Three genes *RYR3, CLCN3*, and *CLCN4* remained and were explored as possible candidates

Exploration of possible candidates from the ion channel gene set

RYR3

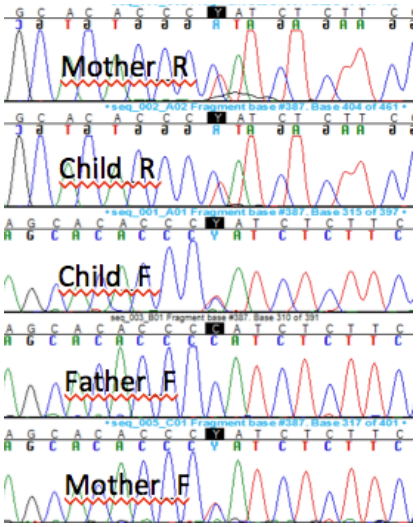
RYR3 (Ryanodine Receptor 3) (ExAC Z score 0.75, RVIS 0.06%) encodes the ryanodine receptor that functions to release calcium from intracellular storage for use in many cellular processes

including a role in triggering muscle contraction. Interestingly, two EE patients with *de novo* *RYR3* variants were identified in the Epi4K study[76]. Here, subject 00505/S2341 has a c.C4135A,p.H1379N *RYR3* variant SIFT 0.21 (tolerated), PP2hvar 0.707 (likely pathogenic), CADD 24.1, gnomad MAF novel. Sanger sequencing studies were carried out but did not validate the variant in the child nor parents.

CLCN3

CLCN3 (Chloride Voltage Gated Channel 3) (ExAC Z score 4.24, RVIS 18.59%) encodes a member of the voltage gated chloride channel family and mediates the exchange of chloride ions against protons. It is expressed primarily in tissues derived from the neuroectoderm and may play an important role in neuronal cell function through regulation of membrane excitability by protein kinase C[237]. Subject 00568/S2297 carries a c.C2216T,p.P739L *CLCN3* variant SIFT 0.08 (tolerated), PP2hvar 0.923 (pathogenic), CADD 27.9, gnomad MAF 7.96 x 10^{-6}. Sanger sequencing studies validated and established inheritance of this variant from his unaffected mother (see Figure 7.20). This variant was classified as a VUS.

Figure 7.20 Chromatogram demonstrating maternal inheritance of *CLCN3* variant in subject 00568.



R denotes sequences obtained by reverse primer and F denotes sequences obtained by forward primer.

CLCN4

CLCN4 (Chloride Voltage Gated Channel 4) (ExAC Z score 4.7, RVIS 7.05%) is an X linked gene that encodes the chloride/hydrogen ion exchanger, which is highly expressed in the brain. Veeramah *et al*. identified a *de novo* variant in a male with EE[230]. Recently, a cohort of 52 males and females with *de novo* and inherited mutations were described with ID, behavioral and seizure disorders ranging from well controlled seizures to EE[240]. Subject 00505/S2341

carries a c.C812A,p.P271Q *CLCN4* variant SIFT 0 (pathogenic), PP2hvar 0.996 (pathogenic), CADD 32, gnomad MAF novel. Sanger sequencing studies carried out did not validate this variant in the child nor parents. This is the same case (subject 00505/S2341) where I was also not able to validate the *RYR3* variant of interest. It is unclear whether a sample mix up occurred here and sample QC processes were re-checked but no error was identified. Re sampling of saliva DNA for this subject was organized and again both variants were not validated.

7.7.3 Monogenic disorders with epilepsy as phenotypic feature gene set

Overall, 37 (mean 0.46) novel CADD>20 monogenic disorder gene set variants were identified in 79 cases compared with 1610 (mean 0.048) out of ExAC_NFE 33370 controls. Gene matching identified 36 genes and per gene chi squared analysis showed 13 genes were significantly enriched ($P<1 \times 10^{-4}$) for novel CADD>20 gene set variants in this MAE cohort compared to the ExAC_NFE cohort. Genes in both groups overlapped and all genes were reviewed and filtered accordingly.

The results of filtering was as follows: 13 genes (*CASR, FKRP, GUCA1, GALNS, GPHN, GPR98*, *LAMA2, MP1, PEX6, PQBP1, SLC17A5*) were de-prioritised due to a negative ExAC Z and/or RVIS score >25th centile, six genes (*CACNA1A, CACNA1H, SCN1A, SPTAN1, SYNGAP1, UBE3A*) were removed as these were known epilepsy associated genes and had been previously reviewed, two genes (*ETFDH, GLRA1*) were de-prioritized due to inadequate nervous system expression, seven genes (*ATR, CNTNAP2, NDUFS1, PLA2G6, PSAP, SAMHD1, SDHA*) were removed due to incongruent gene function or allelic pattern and seven genes (*CEP152, FUCA1, NPC1, PDSS2, PEX1, WDR62, ZIC2*) were deprioritized as these variants were inherited. Only one gene remained following filtering; *PRICKLE2*.

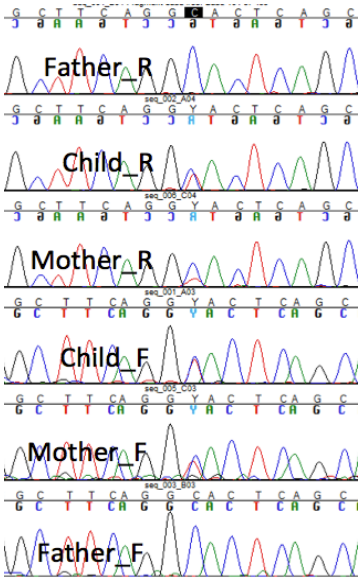
Exploration of possible candidates from the monogenic gene set

PRICKLE2

Prickle Drosophila Homolog of 2 (ExAC Z score 1.51, RVIS 6.78%) is associated with autosomal recessive spinocerebellar ataxia with epilepsy (OMIM 613832). Its role in epilepsy has been debated as the initial pathogenic association with progressive myoclonic epilepsy was subsequently withdrawn[241]. Subject 00530/S2392 has a c.G569A,p.C190Y *PRICKLE2* variant, SIFT 0.41 (tolerated), PP2hvar 0.106 (tolerated), CADD 23.1. Sanger sequencing of the trio

validated the variant in the child and inheritance from his unaffected mother (see Figure 7.21). This gene variant was classified as a VUS.

Figure 7.21. Chromatogram demonstrating maternal inheritance of *PRICKLE2* variant in subject 00530.

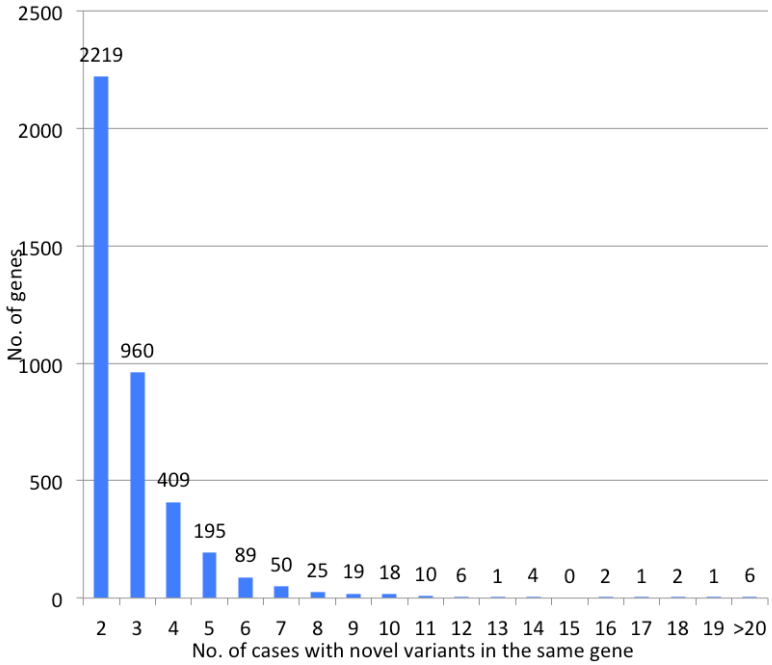


R denotes sequences obtained from reverse primer and F denotes sequences from forward primer.

7.8 Shared novel variant analysis of unsolved cases

Figure 7.22 shows the number of genes with novel coding variants shared between cases. The genes with the most novel variants were *MUC4* (82), *TTN* (50), *AHNAK2* (28), *OBSCN* (26*), IL6R* (24). None of these most commonly shared genes with novel variants were considered good candidates.

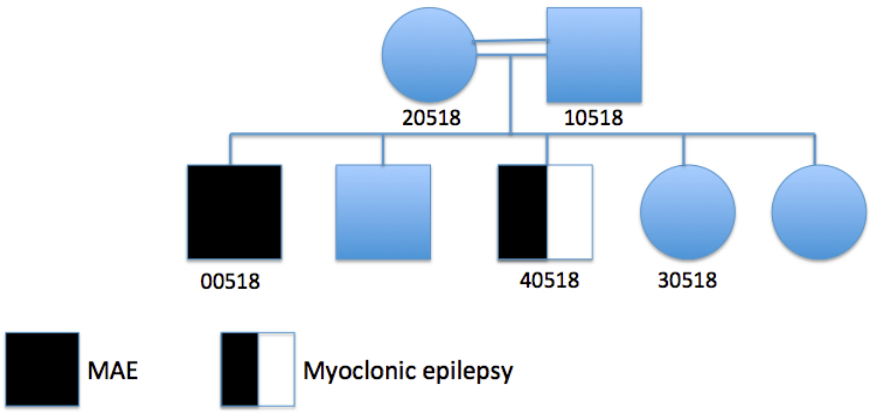
Figure 7.22. Novel gene variants shared between cases.



7.9 Analysis of sibling pair families

Exome sequencing data was available for some sibling pairs with epilepsy. Filtering involved shared rare variant analysis and where sequencing data from other family members were available, segregation of variants was performed. For heterozygous variants, only novel variants were considered and for homozygous variants, a MAF of <0.001 was considered to account for a presumed carrier frequency in the population. Synonymous, non-frameshift and variants annotated with unknown function were removed. Gene variants assembled were reviewed as previously described (see section 6.15.1). PCR and Sanger validation was carried out on candidate genes.

7.9.1 Family one



Family one is a Bangladeshi family with consanguineous parents residing in London. Exome sequencing was performed in subject 00518 and 40518. Subject 00518 has drug resistant MAE. He was a previously normally developing child until his first seizure at 2 years 11 months, which was a GTCS. Subsequent seizure types include myoclonic atonic or atonic seizures and myoclonic seizures. He was recruited at the age of 6 years and at the time he was having daily seizures and was refractory to treatment. On deep phenotyping at 6 years, he has severe ID with cognitive and language age equivalents between 17 to 27 months and atypical autism on 3di. His brother 40518 was recruited at the age of 11 years old, he has a generalised myoclonic epilepsy from early childhood and had been in remission off treatment from 9 years. DNA was collected from other family members (10518, 20518, 30518).

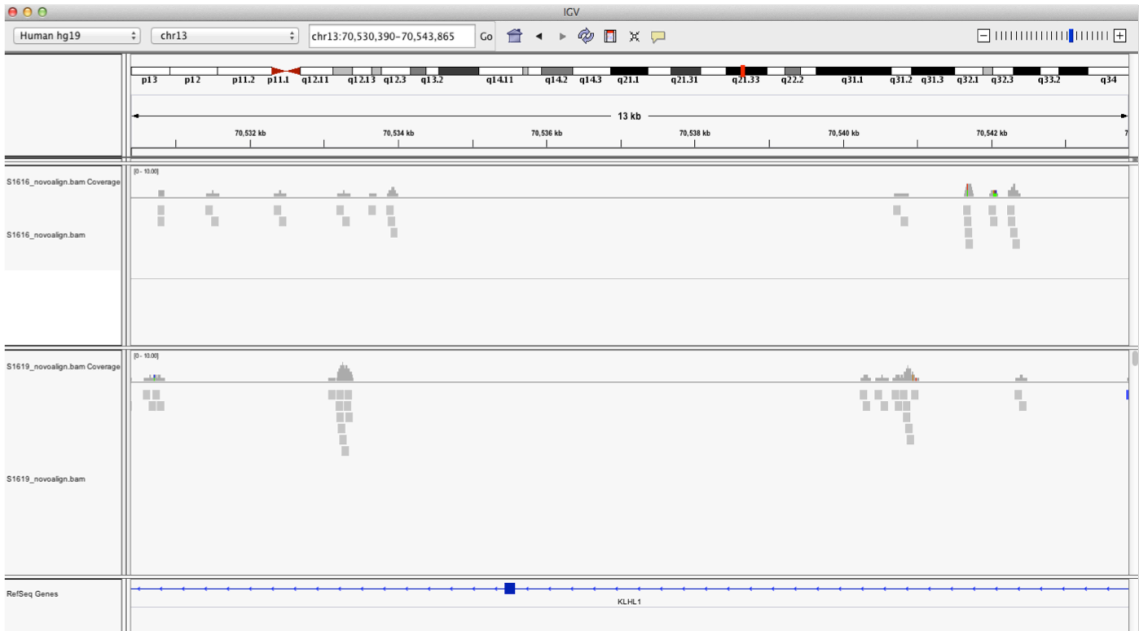
35 shared novel nonsynonymous genetic variants were identified in subject 00518 and his brother 40518. Of note they shared a c.3079G>A:p.A1027T *SPTAN1* variant (ExAC Z score 6.8, RVIS 0.31%) SIFT 0.5 (tolerated), PP2hvar 0.002 (tolerated), CADD 21.2. As described previously, *SPTAN1* associated phenotypes have a hallmark feature of brainstem and cerebellar atrophy and cerebral hypomyelination on neuroimaging. At the time of recruitment, neither child had abnormal neuroimaging findings and the referring clinician was invited to re-review their brain imaging. Additionally all pathogenic variants reported previously cluster in the last two spectrin repeats (19 and 20) in the EF hand domain[225], and the variant described here is located in spectrin repeat 10 and 11. This variant was hence classified as a VUS. No other identified shared novel variants appeared causal based on functions.

As an additional measure, CNV detection using the exome data via Exomedept was explored. This identified a homozygous deletion of 180bp probe within exon 2 of *KLHL1* in subjects 00518 and 40518 (see Table 7.13). This CNV was not in a segmental duplication region. Visualisation of the region on IGV demonstrated an absence of reads from approximately 70533kb to 70541kb covering a single exon in *KLHL1*, exon 2 (see Figure 7.23)

Table 7.13. Exomedept results for subject 00518 and 40518.

Sample ID	Position	Probe	Expected reads	Observed reads
00518	chr13:7053417-70535597	180bp	179	0
40518	chr13:7053417-70535597	180bp	324	0

Figure 7.23 IGV plot illustrating lack of sequencing reads across *KLHL1* exon 2 CNV in subjects 00518 and 40518.

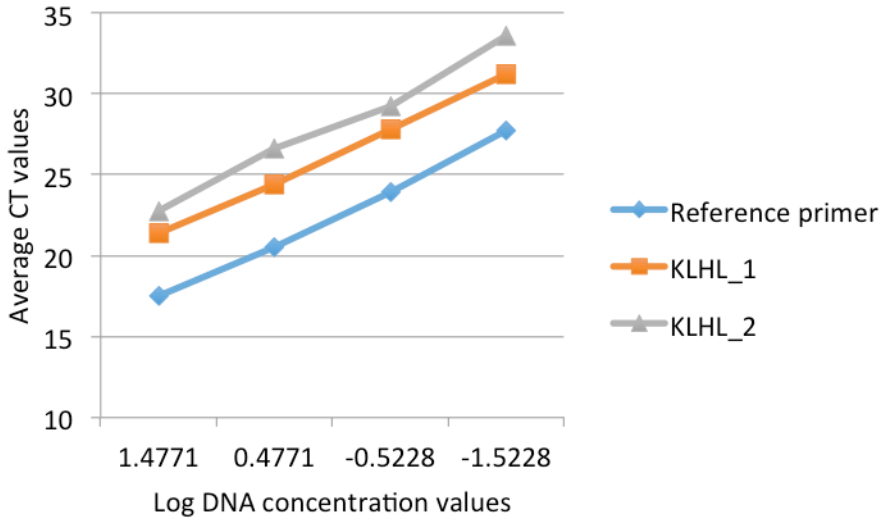


KLHL1

KLHL1 (Kelch-like family member 1) is a neuronal actin binding protein. It is located at 13q21.33 and spans 11 exons. It is primarily expressed in brain tissue and modulates voltage gated CaV2.1 and CaV3.2 calcium channels[242]. *KLHL1* expression was correlated to clinical symptoms of spinocerebellar ataxia type 8 implicating an antisense regulation in neuropathogenesis[243].

Real time PCR was used to validate and identify the CNV within *KLHL1* (see section 6.18.3). The two control samples used were 1052 and 1039. Primer efficiency of KLHL_1 and KLHL_2 primer pairs was checked over a dilution series of DNA (30ng, 3ng, 0.3ng, 30pg) for comparability with the proprietary reference assay (see Figure 7.24).

Figure 7.24. Average CT amounts for primer efficiency on reference DNA.



ΔCT thermocycler triplicate average values are shown in Table 7.14. The ΔΔCT and CN values (copy number inferred value) were then calculated as previously described and shown in Table 7.15. Subjects 10518 and 20518, both had a CN<1, suggesting a heterozygous deletion whereas no product was obtained in subjects 00518, 40518 and 30518 suggesting a homozygous deletion in the region.

Table 7.14. Average CT values at 10ng DNA.

Subject	Reference gene	KLHL_1 primer	ΔCT	KLHL_2 primer	ΔCT
1052	18.9610	23.2464	4.2854	24.9316	5.9706
1039	19.5461	23.5676	4.0215	25.5602	6.0141
00518	20.3937	No product			
10518	19.9083	24.9557	5.0474	26.7301	6.8217
20518	20.3007	25.2925	4.9917	26.8219	6.5212
30518	20.4237	No product			
40518	20.9277	No product			

Table 7.15. CN values for family 518.

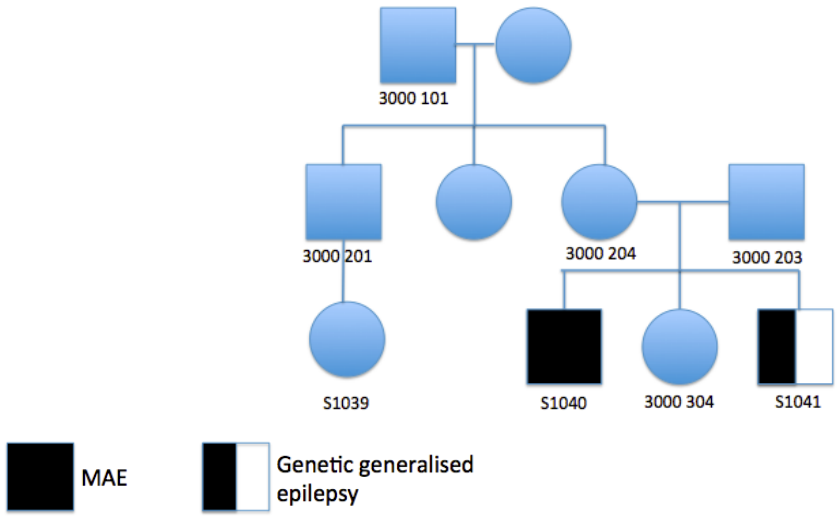
Subject	ΔΔCT1052	CNvs1052	ΔΔCT1039	CNvs1039	Interpretation
00518	No product				Homozygous deletion
10518	0.7619	0.5896	1.0258	0.4911	Heterozygous deletion
	1.7984	0.2874	1.1698	0.4444	
20518	0.7063	0.6128	0.9702	0.5104	Heterozygous deletion
	1.8903	0.2697	1.2617	0.4170	
30518	No product				Homozygous deletion
40518	No product				Homozygous deletion

Upper rows per subject are results for KLHL_1 primer and lower rows are results for KLHL_2 primer, CN copy number inferred value

The two parents are heterozygous for the deletion and all three children, including the unaffected daughter (subject 30518) were homozygous. The variant did not segregate with the phenotype and subject 30518 did not wish to take part in further deep phenotyping studies.

In addition to this, the significance of *KLHL1* copy numbers was explored in other sources. Forty-two Bangladeshi exome sequencing cases recruited for an eczema study, and were available as exome in-house data, were interrogated using Exomedept and no copy number differences were identified involving *KLHL1*. DECIPHER (2015) had 14 entries of mainly large CNVs over the gene, with no epilepsy phenotype.

7.9.2 Family two

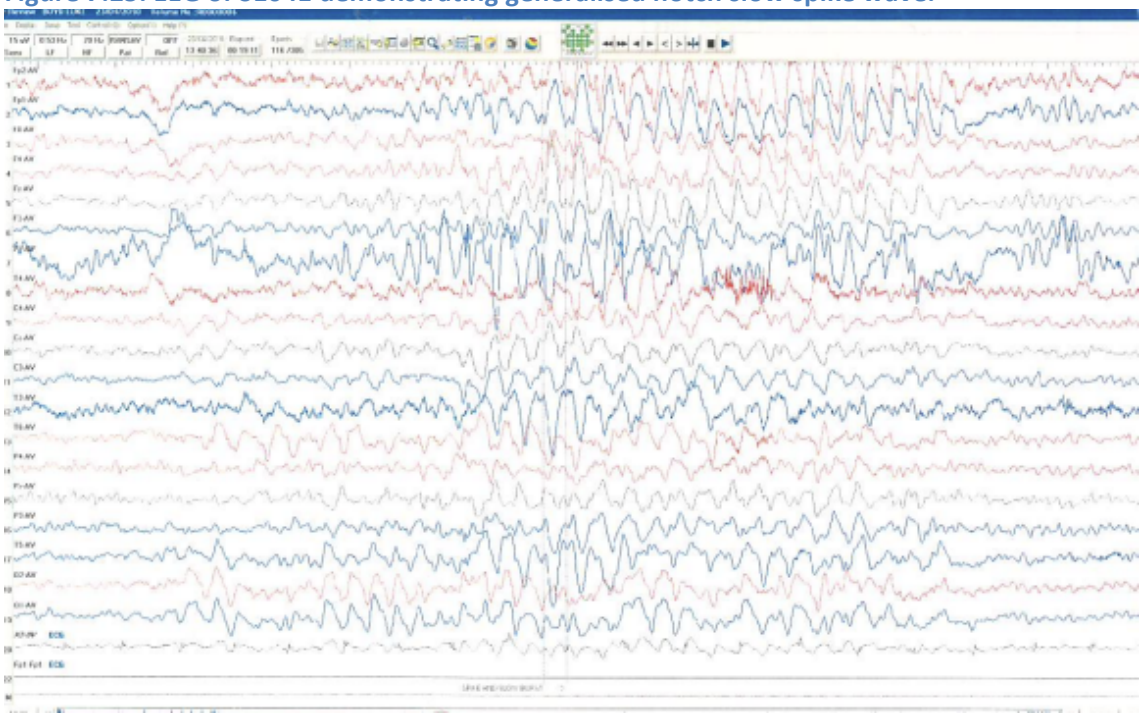


Family two is a White British family with two individuals with epilepsy, S1040 and S1041. The proband S1040 has MAE. He has a background of speech delay prior to the onset of myoclonic atonic seizures which was his only seizure type at 2 years 6 months. His brother S1041 developed GTCS at 10 years and was diagnosed with GGE. He has a history of mild developmental delay and was diagnosed with ASD at 6 years. He also has an intention tremor that is worst in the morning. Whilst their seizure phenotype differs, EEG findings for the brothers are similar. Their EEGs have a normal posterior dominant background with intermittent runs of high voltage notched slow waves which are occipitally dominant and enhanced by eye closure, this sometimes gives rise to generalised bursts of 2-3Hz notched slow waves with spikes, particularly in drowsiness (see Figure 7.25). Their sister 3000 304 had

no history of autism, ID or seizures. She was not available for deep phenotyping. She was classified as unaffected, as were her father 3000 201 and the proband's parents, mother 3000 204 and father 3000 203, and grandfather 3000 101.

Their maternal cousin S1039 is a six-year-old female with a history of blank spells and was deeply phenotyped. Video EEG recording did not capture any clinical or electrographic seizures, even with activation procedures. Neuropsychological assessment achieved a full scale IQ of 87 (19th centile) on the WISC IV. She was classified as unaffected.

Figure 7.25. EEG of S1041 demonstrating generalised notch slow spike wave.



Exome sequencing was performed on S1040, S1041 and S1039. Shared variants in S1040 and S1041 were then searched in their unaffected cousin S1039 and removed if present. 19 shared novel heterozygous variants and 5 homozygous variants were shortlisted. Additional detailed review of individual genes and variants led to selection of a single candidate gene variant *UBE3A*.

UBE3A

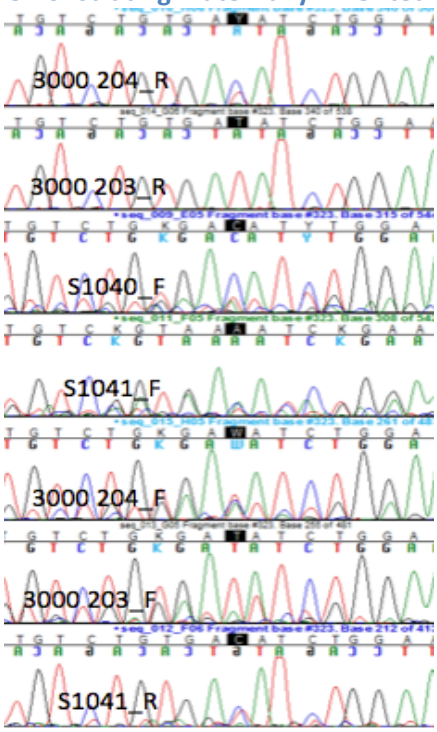
UBE3A (Ubiquitin Protein Ligase E3A) (ExAC Z score 4.2, RVIS 19.54%) is part of the ubiquitin protein degradation system. This imprinted gene is maternally expressed in brain and biallelically expressed in other tissues. *UBE3A* variants are associated with Angelman syndrome (MIM 105830), a neurodevelopmental disorder with severe ID, ataxia, hypotonia, epilepsy and characteristic facies. Four known molecular mechanisms lead to deficient maternal *UBE3A*

expression and Angelman syndrome development: deletion of the Angelman syndrome critical region on the maternal chromosome 15q11.2-q13 (70-75%), paternal uniparental disomy (5-10%), imprinting defects (2-5%) and mutations in the maternal copy of *UBE3A* (10%)[244]. About 90% of patients with Angelman syndrome, carry a *de novo* variant and the rest are a result of *UBE3A* mutations occurring on the maternally expressed chromosome. Loss of functional *UBE3A* gene expression is the principal cause of Angelman syndrome and pathogenic variants are mainly nonsense or frameshift although missense causative variants have been published[245].

Subjects S1040 and S1041 do not have an Angelman syndrome diagnosis but share several Angelman syndrome phenotypic features. Epileptic seizures are common in Angelman syndrome, and 44% of 18 patients in one series[246] and 46.3% of 115 patients in another were described to have myoclonic astatic/atonic seizures. In addition, their EEG feature of generalized slow spike waves has been recognized in patients with Angelman syndrome[247]. The presence of tremor, ASD and ID is also unusual for a GGE phenotype in S1041. The association of *UBE3A* with MAE has been described before by Mefford *et al.* in a single MAE case out of 77 with a *de novo* 15q11 CNV deletion of 270kb encompassing *UBE3A*[90]. Patients with EE and features suggestive of Angelman syndrome have also been reported; a recurrent p.R506C variant in two siblings inherited from an unaffected mother was identified by Carvill *et al.*, as well as a *de novo* p.C604Yfs*23 variant identified by Helbig *et al*.[12,248]. The differing phenotypes in the two siblings in family two has also been similarly described before in a family with a Leu125Stop *UBE3A* variant where the brother had seizures and ID and the sister had Angelman syndrome[249]. Possible reasons for this may be due to unknown genetic, epigenetic or environmental factors, which may modulate the expression of *UBE3A* leading to variable phenotypic expression within the family.

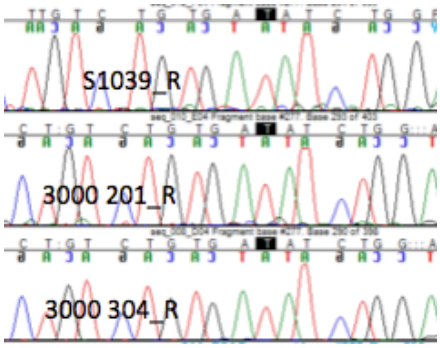
S1040 and S1041 carry a c.A2046G,p.I682M *UBE3A* variant, SIFT 0.01(P), PP2hvar 0.991(P), CADD 24, gnomad MAF novel. There was no match of this variant in 166 entries on Manchester Leiden Open variation database *UBE3A* https://secure.ngrl.org.uk/LOVDv.2.0/ or in 2515 reported cases with Angelman syndrome[245]. Sanger sequencing validated and established that this variant was maternally inherited (see Figure 7.26). Segregation studies with other family members were carried out and specifically their unaffected sister (3000 304) and maternal uncle (3000 201) did not carry the variant (see Figure 7.27).

Figure 7.26. Chromatogram demonstrating maternally inherited *UBE3A* variant in family two.



R denotes sequences obtained by reverse primer and F denotes sequences obtained by forward primer.

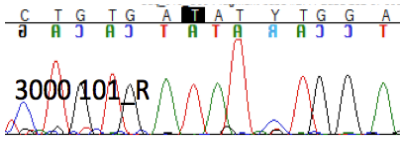
Figure 7.27 Chromatogram demonstrating absence of *UBE3A* variant in asymptomatic family members.



R denotes sequences obtained by reverse primer.

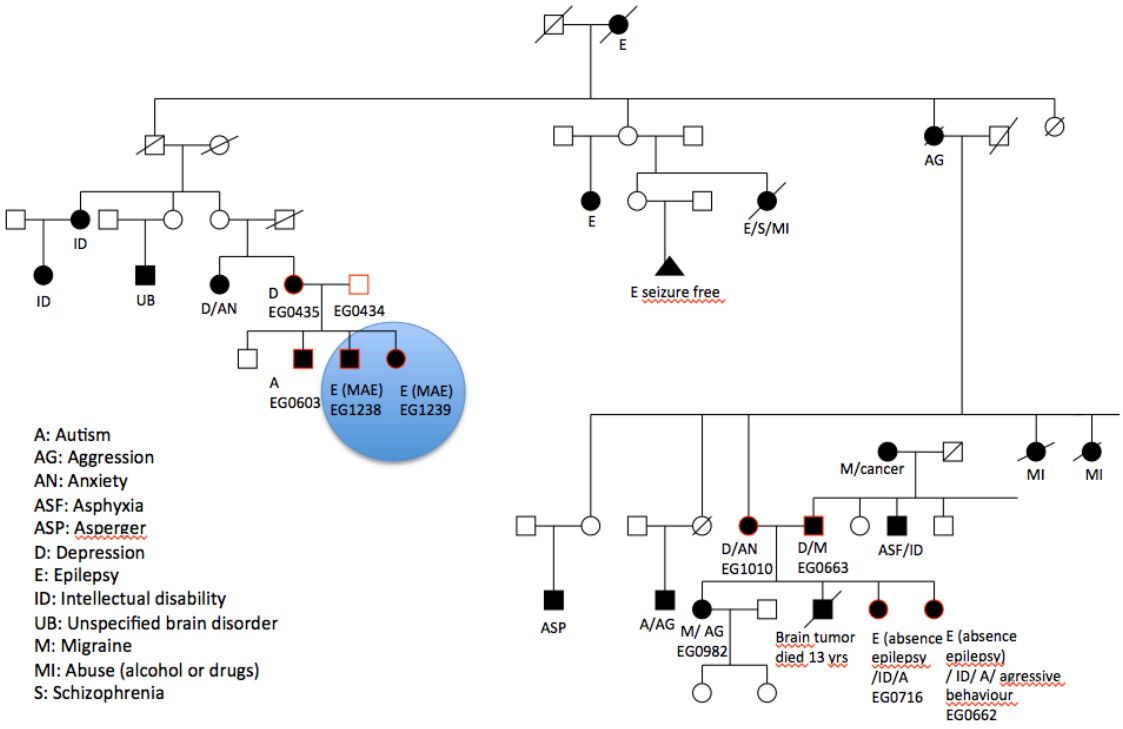
Given the evidence of transmission of the *UBE3A* variant from their asymptomatic mother and absent in their unaffected sibling, the major consideration in a pathogenic hypothesis is the transmission from the maternal grandfather, as the paternal copy of the *UBE3A* is normally silenced through genomic imprinting. DNA from the paternal grandfather was finally collected and he was found not to carry the variant (see Figure 7.28). Therefore, whilst the phenotype and variant characteristics are supportive of pathogenicity, the presence of the variant in their asymptomatic mother, with the variant not paternally inherited in her, cannot be explained and hence this variant was classified as a VUS.

Figure 7.28 Chromatogram demonstrating absence of *UBE3A* variant in maternal grandfather.



R denotes sequences obtained by reverse primer.

7.9.3 Family three



A large multiplex epilepsy family was recruited from the Danish epilepsy centre and exome-sequencing data shared with this study. Two family members had MAE (EG1238 and EG1239), and these cases along with sequencing data from immediate first-degree family members (EG0435, EG0434 and EG0603) were investigated.

Nineteen genes with novel nonsynonymous variants, in 19 different genes were identified in both MAE brothers EG1238 and EG1239. Eight of these genetic variants were also shared with their brother EG0603 with autism. The pedigree suggests that the extended family on the maternal side had a history of neurodevelopmental disorders; hence maternally inherited variants were prioritised. Eight shared variants were maternally inherited and of these three variants from genes *GBP4*, *SLC44A3* and *AKR1C2* were both maternally inherited and shared

amongst subjects EG1238, EG1289 and EG0603. None of these genes appeared to be good candidates on review.

Chapter 8 Discussion

This thesis describes the largest MAE cohort (n=123) assembled to date detailing the main seizure types, EEG features, associated neurological symptoms and neurodevelopmental comorbidities. The spectrum of ID is presented as well as novel data on prevalence of ASD, ADHD, diminished adaptive functioning and other behavioural symptoms in patients with MAE.

EEG studies were performed on 38 first-degree relatives of 13 MAE families. I identified epileptiform activity in two individuals and atypical or benign EEG features in six other individuals. It was not possible to elucidate familial EEG traits or translate to EEG endophenotype studies. I will explore reasons for the differences in my findings with Doose's original cohort and suggest further work, which might aid in clarifying the genetic role of EEG features.

I identified likely pathogenic or candidate variants in 11 of 109 cases through informed filtering methods of exome sequencing data following annotation analysis of previously published epilepsy variants. I identified known genes associated with MAE: a *de novo CHD2* variant (n=1), a *de novo SLC6A1* variant (n=1), a *KIAA2022* variant (n=1) and *SYNGAP1* variants (n=2). Additionally, I identify epilepsy associated genes novel for MAE: a *de novo KCNB1* variant (n=1), a *de novo MECP2* variant (n=1) and a *de novo KCNH5* variant (n=1). I will discuss the selection of three candidate genes: a *de novo SMARCA2* variant in a child with clinical features of both Nicolaides Baraitser syndrome and MAE, a *de novo ASH1L* variant and a *CHD4* variant.

Last, I discuss the genetic heterogeneity of MAE and highlight phenotypic features, which help correlate with known and novel specific gene associations. I propose that MAE is the phenotypic and genetic nosological bridge between genetic generalised epilepsy (GGE) and epileptic encephalopathy (EE). Finally, I explore the applications and future directions in unravelling the complex genetic architecture of MAE.

8.1 Expanding the phenotypic spectrum

8.1.1 The epilepsy phenotype

The general characteristics of this MAE cohort, are remarkably similar to previously published cohorts[2,5,30-34](see Table 1.1). The majority of cases are boys (67.4%) and a family history of

epilepsy was reported in 44 (37.6%) out of 117 cases. The median age of seizure onset is 35 months (range 3 - 72). The characteristic seizure type in MAE is the myoclonic atonic or atonic seizure and this is the most frequent seizure type (93.4%), followed by GTCS (69.1%), myoclonic seizures (68.2%) and absence seizures (54.4%). Focal seizures remain uncommon (6.5%) but are still accounted for and some unusual seizure types such as epileptic spasms (n=4) and myoclonic absence seizures (n=1) are recorded.

Possible explanations for a male bias have been explored before (see Table 1.5), and one plausible reason is the contribution of X-linked genes associated with ID syndromes[250] (e.g. *KIAA2022*, *IQSEC2*), and the recognition of shared genes and pathways between ID and epilepsy[168,170]. The three major seizures; atonic or myoclonic atonic, GTCS and myoclonic seizures are similarly reported in other MAE series[2,31,32,34]. Four cases (subjects 00591, 00598, 00576, 00605) were reported to have epileptic spasms. Epilepsy spasms are characterised by a sudden flexion, extension or mixed flexion-extension of predominantly proximal and truncal muscles, which is more sustained than a myoclonus, but not as sustained as tonic seizures. Epileptic spasms were not reported in earlier cohorts, but were recently described in eight MAE cases by two groups[251,252].

EEG reports were available for 112 cases. Generalised epileptic activity was reported in 104 (92.8%) cases, and additional/isolated focal epileptiform activity was reported in 13 (11.6%) cases. Focal activity has been considered a marker of poor prognosis in MAE[48] but in the MAE cohort here this was not so, as 5/9 cases with focal EEG abnormalities achieved seizure remission (defined as seizure freedom for 2 years), compared to 25/70 cases without focal abnormalities. An abnormal EEG background was reported as a feature in 15 (13.3%) cases within the MAE cohort. These unusual EEG features i.e. focal abnormalities and abnormal background may offer a hint of a monogenic association. For example, subject 291J identified with likely pathogenic *de novo MECP2* variant had an abnormal background with bilateral paroxysmal activity on EEG, additionally she was described as having cerebellar signs which is consistent with a *MECP2* associated phenotype[233] and unusual for MAE.

Cerebellar signs and other abnormal neurological findings were reported in 23 (21.1%) out of 109 cases. The range of abnormal neurological findings in three previous MAE series is 0% to 11.7%[5,31,34]. 13 of the cases reported here had ataxia or tremor. Ataxia was reported in 3/21 MAE cases by Nabbout *et al.* and has been recognised in MAE patients with pathogenic variants in *SLC2A1*, *SLC6A1* and *STX1B*[13,17,20]. Therefore, abnormal neurological signs may be more common then previously reported.

8.1.2 The neurodevelopmental phenotype

Results from the investigation of the neurodevelopmental co-morbidity demonstrated a great degree of difficulty in MAE patients. ID was reported in 76 (64.9%) of 117 cases, ASD in 24 (22.8%) of 105 cases and ADHD symptoms in 39 (41.0%) out of 95 cases. Additionally, high and very high scores in SDQ domains of conduct problems (P<0.0001), hyperactivity/inattention (P<0.0001), peer relationship problems (P<0.0001), and difficulties with prosocial behaviour (P<0.0001) domains were significantly higher than normative populations.

ID based on parental or clinician history was more frequently reported (64.9%) than results from previous published cohorts (mean 45.8%) (see Table 1.2). Formal cognitive testing in 25 cases demonstrated even greater burden of moderate to severe ID (IQ<71) in14 (56%) and four (16%) with mild ID (IQ 71-85). The presence of cognitive difficulties was also evident from results of cognitive and inattentive subscale from the joint CBRS parent and teachers questionnaires, where scores (T score >70) were significantly higher than controls (P=0.0001). This increased prevalence of ID may be due to ascertainment bias from referring clinical collaborators and self-referring parents. Additionally I did not control for multiple factors that may have affected results; such as current seizure burden, antiepileptic drug use and sleep problems. Filippini *et al.* described an MAE case from a group of seven who showed cognitive and behavioural disturbances that remitted entirely with early successful treatment of seizures[53]. There is insufficient evidence to generalise from this one case the effects of treatment on neurodevelopment and a longitudinal cohort study may shed further light. ID is the most important risk factor for ASD in children with epilepsy[106,253] and this is reflected in the burden of ASD symptoms.

ASD symptoms were reported in 22 (21.3%) of 103 MAE cases. This compares to other studies of children with epilepsy where the prevalence of ASD range from 5 to 37%[253]. This broad range reflects the heterogeneity of groups studied and measurement methods in different cohorts. No prevalence rates of ASD in MAE cohorts are available. The UK cohort (61 of the 105 cases) was classified as having autistic symptoms based on parental completion of the social communication questionnaire (SCQ). Comparison of SCQ scores of these 61 cases (43 males, 18 females) with the ALSPAC cohorts showed significantly higher scores in the MAE cohort (P<0.0001) with no male bias (P=0.0901) than were seen in the ALSPAC cohort[116], suggesting that the MAE disease specific effect is greater than the gender effect of autism. When 22 UK cases were tested with the 3di, nine (40.9%) cases fulfilled a diagnosis of ASD. Moreover, 35 (58.3%) out of 63 cases in the UK cohort reported high or very high scores in the prosocial domain of the SDQ questionnaire, which may reflect similar difficulties.

The high prevalence of ASD may be due to a combination of predisposing factors present in MAE: ID, early onset of seizures, and EE[253]. In addition, ASD, ID and epilepsy share causative genes and biological pathways including gene transcription regulation, neurotransmission and maintenance of synaptic structure[166,168,170]. A notable example is *MECP2* related disorders. *MECP2* is a transcriptional activator during brain development and mutations lead to loss of regulation of downstream targets including reduced GABAergic transmission and impaired glutamatergic drive[254]. The phenotype is associated with Rett syndrome and symptoms of ASD, ID and epilepsy[254].

ADHD symptoms were identified in 39 (41.0%) out of 95 cases, and is within the range of 20 – 50% found in groups of paediatric patients with other epilepsies[255,256]. Both parents and teacher's CBRS eliciting relevant symptoms were available in 43 cases. The collection of questionnaires from both teachers and parents explores the pervasive nature of symptoms across these two different settings i.e. school and home. Compared to controls, and correcting for multiple testing in the 10 subscales ($P<0.005$), subscales of DSM IV inattentive domain were significant ($P=0.0001$) whereas the DSM IV hyperactive impulsive domain was not ($P=0.0182$). Several other studies have reported this predominant inattentive presentation rather than hyperactive or impulsive presentation of ADHD in children with epilepsy, leading to more difficulties with complex attention tasks than with simple attention tasks[50,255].

Children with ADHD are more likely to have oppositional defiant disorder, and children with epilepsy and ADHD are more like to have oppositional defiant disorder than children with epilepsy without ADHD[256]. Accordingly, cases with scores in the oppositional domain ($P=0.0005$) of the CBRS and conduct problems domain ($P<0.0001$) in the SDQ questionnaire were significantly higher compared to controls. Additionally 35 (58.3%) of 63 cases sampled report high or very high scores in the impact of difficulties for the child and family SDQ questionnaire. The implication of these co-morbidities can also be seen through the extremely low general adaptive scores.

8.1.3 Adaptive behaviour

This is the first reported study of adaptive behaviour in MAE. I have shown that cases with MAE have significantly diminished adaptive functioning. 41 (69.4%) of 59 cases scored extremely low general adaptive scores on the ABAS questionnaire. General adaptive scores are a composite of the three main domains explored in the ABAS questionnaire: conceptual (e.g. says the names of other people, for example Mama, Daddy or friends' names), social (e.g.

plays with toys, games, or other fun items with other people) and practical (e.g. uses restroom at home without help). By domain, cases scored most difficulties with extremely low scores in practical skills (71.1%), followed by conceptual skills (61.0%) and then social skills (42.3%).

Low adaptive skills have been validated in several studies of children with epilepsy[257-259]. However, the greater difficulties in practical skills in MAE cases warrants further exploration as previous studies of children with epilepsy have indicated greatest difficulties in the communication domain using the Vineland Adaptive Behaviour Scale[257,258]. One possible explanation is the excess of myoclonic atonic and myoclonic seizures as well as cerebellar symptoms in the MAE cases inhibiting motor execution. The presence of ID may be a further contributing factor. Buelow *et al.* investigated IQ and adaptive skills in 50 children with epilepsy and found that correlations ranged from .32 to .45 depending on adaptive skill domain, but mean adaptive scores were more than 2 SD from the norm in both low (IQ<80) and normal IQ (IQ≥80) groups[257], indicating that even in children with normal IQ, an adaptive behaviour screen was warranted.

These findings reveal the impact of cognitive and neurodevelopmental co morbidities in MAE. Children with MAE have increased prevalence of ID, ASD symptoms, ADHD symptoms and other behavioural problems, as well as low adaptive competencies required to function in everyday life. This results in a significant impact on the child and family. Further longitudinal research may offer greater insights. My study highlights the need for initial screening, diagnosis and continued monitoring to help guide intervention for the child at school and at home.

8.1.4 Limitations of phenotype characterisation

The interpretation of the phenotypic characterisation in MAE cases must be taken in context of how the data were collected. Ascertainment bias may have influenced these results. For example, the prevalence of a family history of seizures may have been elevated as a result of how the study was advertised. It is also probable that cases with more severe phenotypes led to greater recall and were referred to the study, resulting in more pronounced estimate of neurodevelopmental comorbidity.

Clinical data collected could also have been improved. Seizure descriptions were based largely on clinical history and/or clinical report. The gold standard for classifying seizure types would be combined video EEG/EMG recordings, although this may not be practically feasible. Next,

EEG features were elicited either from clinician or neurophysiological report. For example, cases recruited from the Euroepinomics cohort completed EEG details through an online platform, which had prefilled fields such as generalised epileptic activity without further elaboration, limiting the quality of data collected. In addition, EEG features evolve throughout the course of the epilepsy and are influenced by co-existing seizure burden and antiepileptic drug use. Moreover, the mainstay of EEG reporting is by expert visual inspection and there is no standardised format in reports, although main neurophysiological features are usually detailed.

Many cases were phenotyped for behavioural and neuropsychiatric symptoms through screening questionnaires e.g. the SCQ and SDQ. These screening tools do not substitute for comprehensive observation and neuropsychological testing, which typically takes several hours and are performed by multidisciplinary teams with expertise across a variety of cognitive and behavioural domains. Nevertheless all the tools used in the study are well-validated instruments and used routinely in clinical practice and research[50,115,116,260,261]. In the cases that were deeply phenotyped, it was sometimes not possible to distinguish the extent of ID due to behavioural difficulties or vice versa. In some instances, the test had to be discontinued (e.g. subjects 00568 and 00596). I also did not explore other environmental and influencing factors such as socio-economic status and sleep difficulties.

8.2 Familial EEG findings: interpretations and limitations

EEG studies were performed on 38 first-degree relatives of 13 MAE families. I identified epileptiform activity in two individuals; subject 20526 was an asymptomatic mother with a right frontal temporal sharp wave, and subject 30574 a 5-year-old sister with 3Hz GSW who was subsequently diagnosed with childhood absence epilepsy. It was not possible to draw further conclusions regarding inheritance of familial EEG traits or extend into the use of EEG traits as an endophenotype due to the paucity of positive findings.

However, there was an excess of epileptiform features among relatives when divided according to age and compared to control populations of >16 years[124] (P=0.05, RR 6.82) and <16 years[127] (P=0.58, RR 1.69), leaning towards statistical significance in adults but not in children. There were also no statistical differences in prevalence of PPR and GSW/sharp waves in the MAE cohort here (2/38, 5.2%), when compared with Doose *et al.*'s cohort (33/151, 21.8%)[5], although these P values may not be accurate due to the small sample sizes. Consequently, RR analysis in Doose *et al.*'s cohort compared with this MAE family cohort

demonstrated a higher prevalence for PPR (RR 3.71 in siblings, RR 5.61 in parents) but not so for GSW or sharp wave (RR 0.36 in siblings, RR 1.08 in parents). In sum, parents/adults had an excess of epileptiform features compared to controls and PPR were underreported compared to Doose *et al*.'s cohort.

The identification of epileptiform features in EEG studies of first-degree relatives have been recognised in other epilepsies[63,262]. Pal *et al*. reported 2.3% of first-degree relatives with GSW in 139 probands with various sub syndromes of GGE[63], while Jayalakshmi *et al* identified GSW in 9.8% of first-degree relatives in 31 probands with JME[262]. The increased identification by Jayalakshmi *et al*. may be due to 20% consanguinity in their cohort and a sleep deprivation EEG adding to familial susceptibility and seizure predisposition in their cohort.

Several factors may account for the lack of identifiable epileptiform features in siblings (<16 years) and PPR in both parents and children. First, the small sample size particularly in the sibling group in this MAE cohort, and incomplete families recruited limits the power of discovery. It is also now acknowledged that Doose's original cohort consist of a heterogenous group of epilepsies and may not be a suitable direct comparison cohort. Next, recording conditions were vastly different. Doose carried out paper EEGs with 8 electrodes whereas all EEGs performed here were on digital recordings with electrodes laid out in the 10-20 system. Additionally, it is unclear whether Doose carried out sleep recordings, which were obtained here, and the lack of activation procedures in some study subjects here would have compromised results, particularly PPR findings. The definitions in EEG abnormalities may have differed; dysrhythmias, which were defined by Doose as irregular background activity with widely varying amplitudes and frequency, such that no dominant activity is recognisable, was not accounted for in my study.

In an effort to address the significance of dysrhythmias, quantitative EEG power spectral studies of the five-frequency bands, delta, theta, alpha, beta and gamma were performed. Four family members had to be eliminated due to difficulties in obtaining a suitable artefact free epoch. Comparison in the five frequency bands demonstrated differences in theta and beta bands in parents versus siblings with no differences in mothers versus fathers. These differences could be accounted for by physiological EEG differences with increased theta frequencies in siblings due to their younger age and more enhanced beta rhythm in adults. Comparison with control data from other investigators was not possible due to differences in how the EEG data were collected. Altogether to improve upon this arm of the project, EEG recordings carried out to meet quantitative EEG analysis criteria for both family members and

age matched controls, along with activation studies in all subjects, sleep recordings and a larger cohort would deliver a clearer picture of the heritability of specific frequency bands and EEG features.

Another unexplored strategy akin to the genetic linkage studies performed in PPR/GGE families[66,67], would be to stratify MAE families investigated by only performing EEG recordings in MAE families with multiple affected members with epilepsy. Indeed I had recruited 16 families with a first-degree relative with epilepsy but was only able to perform family EEG studies in three of these families (family 518 (incomplete), 574 and 602). This was because most families were located away from the principal study site in London, UK. The advent of home video EEG and/or portable EEG recording[263] may be a useful aid in capturing a cohort of rare multiplex MAE families for dissection of familial EEG traits.

8.3 Expanding the genetic spectrum of MAE

I assembled an exome-sequencing cohort of 109 MAE cases. I identified likely pathogenic or candidate variants in 11 of 109 cases. This comprised known genes associated with MAE: a *de novo CHD2* variant (n=1), a *de novo SLC6A1* variant (n=1), a *KIAA2022* variant (n=1) and *SYNGAP1* variants (n=2); epilepsy associated genes novel for MAE: a *de novo KCNB1* variant (n=1), a *de novo MECP2* variant (n=1) and a *de novo KCNH5* variant (n=1); and three candidate gene variants: *SMARCA2, ASH1L and CHD4*. Annotation analysis of known epilepsy variants assisted in providing guidance for subsequent filtering of variants.

8.3.1 Annotation analysis of known epilepsy variants

Annotation analysis of known epilepsy variants revealed useful insights for further analysis into the MAF and prediction capabilities of *in silico* predictors. Previously published epilepsy variants were curated and segregated into two groups, (1) the All_epilepsy group which consist of heterogenous disease associated epilepsy variants and (2) the Pure_epilepsy group which consist of variants associated with specific epilepsy syndromes. In the All_epilepsy group, 83.7% of variants were novel (not in ExAC, 1000G and ESP) and 95% were ultra rare with a MAF<0.0001. In the Pure_epilepsy group, 97.2% of variants were novel. Novelty was further validated in my cohort when all pathogenic variants identified by the Euroepinomics consortium (see table 7.4) and by my filtering (see table 7.9) were novel even when cross-referenced with gnomAD, a recently launched population database of 140 000 exomes and

genomes, indicating that novelty is the most consistent characteristic for a MAE genetic variant. *In silico* predictors were less consistent.

In silico prediction using CADD, SIFT and PP2hvar predicted 51.2% of variants to be pathogenic in the All_epilepsy group and 64.8% of the Pure_epilepsy group. The use of CADD alone proved to be a better predictor. A CADD score of 20 captured 93% of Pure_epilepsy variants, whereas SIFT predicted pathogenicity in 82.4% and PP2hvar predicted pathogenicity in 80.4%. CADD scores as a measure of deleteriousness has been validated before in epilepsy studies. CADD scores within EE genes were assessed in 360 patients with epilepsy, patients were more likely to have novel and high CADD score variants relative to the ExAC population[150]. As a consequence to this annotation analysis, I used novelty and CADD as a prioritization tool during gene exploration of aetiologically relevant gene sets.

8.3.2 Identification of known MAE genes

I identified five cases with variants in four different known genes associated with MAE. A *de novo CHD2* variant (n=1), a *de novo SLC6A1* variant (n=1), a *SYNGAP*1 variant (n=2) and a *KIAA2022* variant (n=1); the latter two genes had variants of unknown inheritance in two subjects. These variants were classified as likely pathogenic due to supporting phenotype genotype correlation, and add to literature of the phenotypic association of each gene.

Distinctive phenotypic features not reported in the rest of the MAE cohort, and correlating to specific genetic associations, were identified in two cases. Subject 00559 has a *de novo* c.3734+7A>G *CHD2* variant. *De novo CHD2* variants have been previously reported in four other MAE cases[12,14,15] and are also associated with several other epilepsy phenotypes such as SMEI and LGS[12]. Nevertheless, a unique phenotypic feature identified by Thomas *et al*. in *CHD2* associated epilepsy is the atonic myoclonic absence seizure[15], this distinctive seizure type was reported in subject 00559. Including the MAE cohort here (1/67 cases), the adjusted genetic contribution of *CHD2* variants in MAE reduces from 2.9% to 2.3% (see section 1.8 for genetic contributions). A *de novo* c.C1155G,p.F385L variant was also identified in the recently identified MAE gene *SLC6A1* adjusting its genetic contribution from 3.7% to 3%, based on the cohort described here. Seven previous MAE cases have been reported with *SLC6A1* variants[17,75]. Subject 00595 had an unusual EEG feature of posterior eye closure sensitivity which was also reported in another *SLC6A1* patient[17], illuminating this feature as a phenotypic clue in *SLC6A1* associated epilepsy.

The above two cases have *de novo* variants, now an established pathogenic principle in EE and neurodevelopmental disorders, largely due to the success of trio based sequencing studies[76,155,165,230]. The human mutation rate per nucleotide per generation is in the order of 1.18×10^{-8} (SD $\pm 0.15 \times 10^{-8}$)[264]. Therefore, each individual is expected to have ~ 1 *de novo* variant in their exome and a significant excess of *de novo* variants with an average of 1.2 variants were identified in 356 EE cases compared to the general population ($P=8.2 \times 10^{-4}$)[155]. *De novo* variants in neurodevelopmental disorders tend to be more prevalent[165,171,172], and are more enriched for gene disrupting mutations (nonsense, splice site and frameshift)[165,172,174]. *De novo* variants are regarded as more deleterious with a stronger disruptive effect on biological functions due to less stringent evolutionary selection[265]. Pathogenic *de novo* variants are thought to account for at least 12% of EE[155] and 42% of developmental disorders[165]. The role of *de novo* variants in selected genes is undisputed in MAE. Nonetheless, even without inheritance studies, it is possible to make a judgement on pathogenicity.

Subject 138J carries a p.L421fs *KIAA2022* deletion and subject 00514 carries a p.8854_860del. *SYNGAP1* deletion. Inheritance studies were not possible due to lack of parental DNA. However these variants were classified as likely pathogenic due to the consistent phenotype and variant characteristics with previously reported pathogenic variants. *KIAA2022* is an X linked gene first identified in males with ID[266]. Epilepsy and ID are common associated features with *KIAA2022* mutations but only one female patient with MAE has been published[23]. Reported pathogenic variants are all frameshift or premature stop codon and mostly located in exon 3[23]. Subject 138J has ID and her *KIAA2022* variant carries similar features to reported pathogenic variants and therefore was classified as likely pathogenic, recognising subject 138J as the second female patient with MAE to carry a *KIAA2022* mutation. X-inactivation and RNA expression studies were not performed.

Subject 00512 and 00514 both carry a *SYNGAP1* variant with features that fit with *SYNGAP1* pathogenic variants, they both have frameshift deletions located within exons 8 to 15[16]. Additionally they both have a very severe phenotype, with pre-existing global developmental delay and ataxia, which is consistent with *SYNGAP1* encephalopathy. These two cases here adds to the three other reported MAE cases which are associated with *SYNGAP1*[16], elevating *SYNGAP1* as one of the more prevalent genes for MAE.

While *SYNGAP1* variants are more recognised in MAE, the lack of *SLC2A1* variants identified here, which was previously reported in 4/84 MAE patients[13] is intriguing. *SLC2A1* variants are well recognised to be associated with glucose transporter 1 deficiency syndrome, GGE and

early onset absence epilepsy[18,267]. Along with European colleagues, we identified no *SLC2A1* variants in 120 MAE cases[18]. Updated with further subjects from this cohort, no *SLC2A1* variants are identified in 187 MAE cases. Reviewing the previously described four MAE patients with *SLC2A1* mutations; two patients had regular paroxysmal exertional dyskinesia, one patient had ataxia, dysarthria, poor motor skills and deceleration of head growth and one patient was part of a larger family with absence seizures and paroxysmal exertional dyskinesia[13]. All these patients therefore have overlapping phenotypic features with glucose transporter 1 deficiency syndrome and suggest that MAE may be a rare expression of *SLC2A1* mutations.

Other genes associated with MAE, and which were described in the introduction, were identified in this cohort by the Euroepniomics consortium (*KCNA2*[19], *SCN1A*, *STX1B*[20]) and through other concurrent studies (*TBC1D24*[22]). Next I will discuss genes that have not been previously associated with MAE and with variants that I have classified as likely pathogenic.

8.3.3 Identification of novel MAE genes

I expand the genetic spectrum of MAE by identifying epilepsy associated genes novel for MAE. These are a *de novo KCNB1* variant (n=1), a *de novo MECP2* variant (n=1) and a *de novo KCNH5* variant (n=1). Individually these genes have distinct yet overlapping phenotypic features to account for an MAE phenotype.

A recurrent *de novo KCNB1* p.R306C variant was identified in subject 00533. *KCNB1* variants are generally associated with an infantile epilepsy/spasms phenotype with severe ID, and seven other patients with pathogenic *KCNB1* variants including one patient with a p.R306C have been reported[101,228,229,268]. Subject 00533 has a severe MAE phenotype with seizure onset at 6 months. Seizure onset of 6 months is at the lower limit for MAE based on the ILAE definition[29], and therefore *KCNB1* associated MAE can be best considered with an early onset severe phenotype. This p.R306C variant occurs at a conserved positively charged residue in the S4 segment in the human Kv2.1 protein which is important for voltage sensing[269]. Saitsu *et al.* performed whole cell patch clamp technique in Neuro2a cells and primary cortical neurons and showed that the p.R306C mutation reduced sensitivity and responsiveness of the voltage sensor for channel opening with inhibited repetitive neuronal firing. The authors postulated that insufficient firing of pyramidal neurons would disturb both development and stability of neuronal circuits leading to disease phenotypes[229]. Another member of the voltage gated potassium channel gene family was also identified with a likely pathogenic variant.

Subject 00525 was found to have a *de novo KCNH5* splice site variant c.1569+7G>T. Only one other patient with epilepsy with a *de novo* p.R327H *KCNH5* variant has been published[230]. *KCNH5* is expressed predominantly in the brain[270] and the authors considered this gene variant a good candidate and subsequently performed voltage clamp analysis of the mutated protein. This showed a strong hyperpolarisation shift of voltage dependent activation and they hypothesized that the R327H mutation weakens ionic interactions and favour voltage gated potassium channel Kv10.2 opening[231]. That patient and subject 00525 share phenotypic features of a severe epilepsy phenotype associated with ID (EE). Additionally their EEGs demonstrated focal spikes, which is an unusual feature for MAE. As for phenotypic differences, the p.R327H patient had seizure onset at 6 months and did not have myoclonic atonic seizures whereas subject 00525 had seizure onset at 3 years and myoclonic atonic seizures; in keeping with her classification of MAE. This c.1569+7G>T variant is located in the cytoplasmic domain outside the S0-S4 helical pore. Blood RNA samples for subject 00525 have been collected to investigate whether her splice variant has effect on expression. An update of her clinical phenotype revealed that she is now seizure free on a medium chain triglyceride diet, and is off antiepileptic medication.

The accessibility of genetic data in the NGS era has led to an expansion of related phenotypes associated with genes previously thought to be only accountable for specific congenital syndromes. This is so for the next gene to be discussed *MECP2*, which were identified in one MAE case. These genes are quietly moving on from syndrome genes to epilepsy-associated genes as well often with suggestive syndrome features.

The phenotypic spectrum associated with *MECP2* is classic Rett syndrome (OMIM 3127250), variant Rett syndrome, neonatal encephalopathy and X linked ID[233]. The most well recognised association is classic Rett syndrome, which is a neurodevelopmental disorder characterised by developmental regression, stereotypical hand movements, motor abnormalities and epilepsy mainly in females. Seizures are present in up to 94% of individuals with classic Rett syndrome, and a slow background with loss of occipital dominant rhythms on EEG is observed[232]. Subject 291J is an Italian female who carries a *de novo MECP2* p.P225T variant. She has overlapping phenotypic features found in classic Rett syndrome of cerebellar signs and an EEG with a disorganised background. *De novo MECP2* variants have been recognised as causative for patients with epilepsy by other investigators through epilepsy gene panels[248,268]. Parrini *et al*. identified *MECP2* variants in 4/349 patients with epilepsy; all patients identified had drug resistant epilepsy with seizure onset from 22 months to 5 years. The authors initially

recognised two patients with Rett syndrome, but subsequently reclassified all cases to have features consistent with Rett syndrome, including one patient who progressed during adolescence to a PPM-X phenotype of ID, parkinsonism, pyramidal signs and parkinsonism (PPM-X)[268]. Helbig *et al*. identified *MECP2* variants in 4/293 patients with epilepsy, two patients were classified with Rett syndrome but the others were reported as an EE in one and unclassified epilepsy in the other[248]. Including subject 291J, these cases support a causative role of *MECP2* in epilepsy, often with overlapping clinical features of Rett syndrome.

Mouse models have indicated that MeCP2 regulates the expression of a wide range of genes in the hypothalamus through both transcriptional activation and repression[254]. This p.P225T variant is located in the transcription repression domain (TRD), a well characterised canonical functional domain in MeCP2 and, where many pathogenic variants present. The TRD domain is part of the co-repressor interacting region and is involved in transcriptional regulation by interacting with the corepressor Sin3A and histone deacetylases[271].

A potential clinical implication for subject 291J is that children with Rett syndrome have an increased risk of life threatening arrhythmias associated with prolong QT_c interval and should avoid drugs known to prolong the QT_c interval[235]. It is unknown whether this also applies to patients with a *MECP2* associated epilepsy and the natural history and complications of these patients remain to be described.

The genes described in this section were identified due to their known association with epilepsy. Next, I will describe genes without known association with epilepsy that I have identified as MAE candidate genes.

8.3.4 Discussion of MAE candidate genes

Three genes were identified as candidate genes in this cohort. These were in subject 3003 301 with a *de novo SMARCA2* variant[52], subject 00530 with a *de novo ASH1L* variant and subject 00526 with a *CHD4* variant.

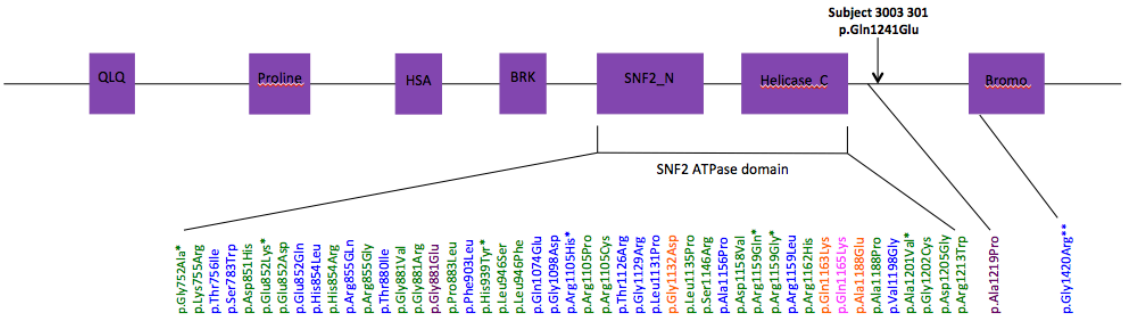
Subject 3003 301 with MAE, ASD and ID, was identified with a *de novo* p.Q1241E *SMARCA2* variant which led to clinical re-examination and a diagnosis of Nicolaides Baraitser syndrome (NCBRS). NCBRS may not have been recognised earlier because (1) the NCBRS phenotype is progressive and in younger ages the features are subtle and difficult to recognize, (2) in NCBRS,

neither seizure semiology nor specific epilepsy syndromes have been well described, and (3) there is no reported association between MAE and NCBRS.

Epilepsy occurs in two-thirds of patients with NCBRS, and in a series of 61 individuals with NCBRS, the median age at the first seizure was 18 months[182]. Generalised seizures were common in a series of 11 patients with NCBRS, ten had GTCS and six had atonic seizures (ECE 2016 abstract, Hofmeister *et al*.). Additionally, as in our patient, a co-occurrence of decreasing mental abilities with the onset of seizures (EE) in NCBRS has been observed[182]. EE is also frequent in MAE, and it is possible that the co-existence of NCBRS /MAE is more common than expected.

At least 80% of patients with NCBRS carry a variant in *SMARCA2* (OMIM 601358) but to our knowledge all point mutations have been located in exons 15 - 25, and never before in exon 26[181]. *SMARCA2* is located on chromosome 9p24.3 and its longest transcript has 34 exons. Thus far, 62 missense mutations and 3 in-frame deletions clustering in the ATPase domains of exons 15 - 25 of *SMARCA2* have been reported in patients with NCBRS (see Figure 8.1 for schematic diagram of SMARCA2 protein and location of variants)[182,272,273]. These pathogenic variants are thought to have a dominant negative effect by abolishing the ATP hydrolyzing engine potential of the protein, removing the ability of the protein to reposition histones on DNA.

Figure 8.1 Schematic diagram of SMARCA2 protein and its domains and location of missense mutations associated with NCBRS.



QLQ glutamine-leucine-glutaimine domain, Proline proline rich domain, HSA small helicase/SANT associated domain, BRK brahma and kismet domain, SNF2_N and Helicase C, Bromo acetyl lysine binding domain. Patient's variant in black. Variants first reported by Van Houdt *et al*. in green[181], Wolff *et al.* in orange[274], Sousa and Hennekam in blue[182], Bramswig *et al.* in purple[272] and Ejaz *et al.* in pink[273]. * recurrent mutations, ** patient without classical NCBRS but with overlap features of ID, absent speech and seizures.

Recently, one other patient with NCBRS but no seizures has been reported to carry a mutation p.A1219P, located outside but close to the ATPase domain[272]. To our knowledge, the patient

reported here carries the first mutation located in exon 26, also in close proximity to the ATPase domain. We hypothesize that both mutations similarly affect ATP hydrolysis and chromatin remodelling ability. Sousa and colleagues reported a case with a mutation in the Bromo domain, in exon 30 (p.G420R)[182]. Their case did not have the typical facial features seen in NCBRS but had overlap features of severe ID, absent speech and seizures. The absence of variants reported outside exons 15 - 25 may be due to restricted diagnostic laboratory sequencing excluding other areas of the gene based on previous association[182]. Based on our findings, we would suggest incorporating exon 26 in clinical laboratory testing.

Genetic aberrations involving exon 26 in addition to ATPase disruption have also been described. NCBRS has been reported with a 32 kb *de novo* in-frame *SMARCA2* deletion affecting exons 20 to 26 and a 55kb interstitial deletion involving exons 20 to 27 was reported in a patient with the phenotypically similar Coffin-Siris syndrome[274,275]. These patients had epilepsy, but the associated seizure semiology and epilepsy syndromes merit further description.

SMARCA2 is one of the six genes that encodes the catalytic subunit components of the SWItch/sucrose nonfermentable like chromatin-remodeling complex (SWI/SNF complex). The SWI/SNF complex is evolutionarily highly conserved from yeast to humans and works as a chromatin remodeler; altering chromatin structure through ATP hydrolysis. SWI/SNF proteins regulate gene expression by re-positioning nucleosomes and altering DNA transcription. A close connection between the SWI/SNF complex and neurological development has been implicated through the identification of numerous variants in genes encoding the subunits in a range of neurodevelopmental disorders such as Coffin-Siris syndrome, sporadic ID, ASD, schizophrenia and Kleefstra syndrome[276].

In conclusion, the epilepsy phenotype in NBCRS can be consistent with MAE syndrome, and NBCRS spectrum features can be caused by variants outside the ATPase region of *SMARCA2.* Like *SMARCA2*, the next gene identified as a candidate*, CHD4* has a role in chromatin remodelling.

Subject 00526 is a 6-year-old female with severe MAE and ID. Her parents are first cousins and have no history of epilepsy. Her mother (subject 20526) was identified with an incidental epileptic sharp wave in familial EEG studies performed here. Subject 00526 has a novel p.H896R *CHD4* variant with supportive *in silico* predictors. The curated gene within the neuropsychiatric gene set which led to identification of this gene variant was a *de novo CHD4*

variant in a single patient with EE[101] and two patients from the DDD study who did not have seizures[162]. Lately, five individuals with *de novo* missense substitutions in *CHD4* were reported with overlapping phenotypes of developmental delay, hearing loss, macrocephaly, distinct facial dysmorphisms, palatal abnormalities, ventriculomegaly and hypogonadism. A further four patients were identified in the updated DDD study, but they had no history of seizures[165,238].

CHD4 encodes the Chromodomain helicase DNA-binding protein 4 (CHD4), which is an ATP dependent chromatin remodeller like SMARCA2. Chromatin remodelling is an epigenetic mechanism that controls DNA accessibility to transcription, replication and DNA repair regulation. CHD4 helps drives this process by mobilising nucleosomes and modifying DNA packaging[277]. In addition, *CHD4* belongs to the CHD subfamily II, and among its paralogs is *CHD2*, which is associated with MAE[12,14,15]. The variant identified in subject 00526 is located within the ATPase/helicase domain (http://www.uniprot.org), and may disrupt the ATPase activity of CHD4. In mouse models, Chd4 is highly expressed in the central nervous system (http://www.informatics.jax.org) and the lack of Chd4 results in impaired parallel fiber/Purkinje cell synaptic connectivity in the developing central nervous system[278].

The variant identified in *CHD4* summarises as a suitable candidate gene: RVIS 2.82%, ExAC Z 7.05, gnomad MAF novel, predicted pathogenic by SIFT and PP2hvar, CADD 24.6. The spectrum of phenotypic expressivity of seizure-associated genes have been suggested in the updated DDD study; where 285 of 4293 individuals had truncating or missense *de novo* variants in known seizure associated genes but 56% of these individuals had no coded terms related to epilepsy[165]. Unfortunately I was not able to perform validation or inheritance studies in this case and hence I have classified this variant as a VUS. However it remains a good candidate for further studies to better understand the phenotypic variability of *CHD4*.

Gene exploration with the neuropsychiatric gene set identified another candidate gene, *ASH1L.* Subject S2392/00530 is a 7-year-old male with seizure onset at 6 months. He has severe MAE with multiple refractory seizure types. He has moderate to severe ID, autism and ADHD. There is no family history of epilepsy. He has a *de novo* p.R1342X *ASH1L* variant. The curated variant within the neuropsychiatric gene set was a *de novo* p.A724S variant identified in 1/100 patients with ID. The patient reported by de Ligt *et al.* is a 37 year old male with ID, obsessive behaviour with sand and water, rotating objects, and has facial dysmorphisms of deeply set eyes, synophrys, small asymmetric ears, a beak shaped nose, high nasal bridge and short philtrum. He also has pectus carinatum with wide spaced nipples[175]. Two other *de novo*

ASH1L variants have been reported in ASD patients, detailed phenotype in one reported no history of seizures[279,280].

ASH1L (Absent, small, or homeotic disc 1 Like Histone Lysine Methyltransferase) encodes a member of the trithorax group of histone modifiers, histone methyltransferase which is involved in gene activation. The expression of Ash1L is enriched in the brain and a role in epigenetic modification in brain functioning was implicated when Ash1L knockout mice completely abolished the activity dependent repression of neurexin 1α, a presynaptic adhesion molecule required for synaptic formation[236]. This highly conserved variant is located between two DNA binding regions A.T hook 1 and 2, and may result in nonsense mediated decay reducing the production of full length functional protein and affecting epigenetic modification in regulating brain function.

Summing up the evidence, subject S2393/00530's phenotype is supportive given his neurodevelopmental features of ID and ASD which are present in other patients with *de novo* *ASH1L* variants. It is difficult to extrapolate a direct causative association for *ASH1L* in MAE although this variant is considered a good candidate and at least partly causative for the subject's phenotype.

8.3.5 Limitations of genetic analysis

The main limitations of the genetic analysis can be divided into the following: (1) limitations in exome sequencing, (2) poor detection of somatic variants, (3) use of novelty as a filtering criteria and population databases, (4) limitations of *in silico* predictors, (5) lack of molecular validation and segregation studies in some cases, and (6) lack of functional work.

Limitations of exome sequencing have been discussed in section 6.6. Exome sequencing is poor in identification of structural variants such as translocations and inversions, indels, copy number variants and triplet repeats. This technique does not explore potentially disease-causing variants in intronic regions, promotor, enhancer regions and micro RNAs. In addition, false positive and false negative calls can occur due to errors in sequence generation and alignment. Some of these calling errors were identified during IGV interrogation of candidate variants.

IGV interrogation and classification of all selected variants was performed in order to identify calling errors. Indeed several variant calls were identified to be the result of errors in

alignment or sequencing, such as the same indel identified in multiple samples in *SIK1* (see Table 7.6). This classification of calling of variants may have dismissed somatic mosaic variants. Specifically, I considered variant calls in less than 20% of total reads for heterozygous variants as failed, although of variants classified as failed, no variants were classified as failed based on this criterion. Somatic mosaic variants are recognised in genes related to brain malformations and epilepsy such as *MTOR* and *PIK3R2*[217,220,221], which were reviewed here. Hence, I may not have identified variants that were present in other tissues such as the central nervous system.

Novelty was used as a filtering criterion. This was defined as variants not identified in 1000G, ExAC and ESP, and subsequently a gnomAD MAF was annotated following its launch. It is worth considering potential confounders within these population databases. First, these databases may not account sufficiently for ethnic populations such as Middle Eastern and African populations and may lead to under representation of MAF in these cases[160]. Second, the individuals ascertained for these databases had biomedically important adult onset diseases, and to that end, it is widely understood that that these control databases contain disease associated variants[281], although severe paediatric diseases were excluded. Third, the re-annotation of a gnomAD MAF on variants that were previously annotated novel reveals the scope of diversity of protein coding genetic variation that remains to be investigated. Nevertheless, all published pathogenic variants identified in the Euroepinomics cohort were novel in all four population databases (see Table 7.4).

I used a CADD score of >20 as a filtering parameter during burden analysis of aetiologically relevant gene sets. CADD is a high performing pathogenic predictor that captured 93% of Pure_epilepsy variants using a cut off score of 20 (see section 7.3.2). CADD has also been verified by other investigators in filtering for epilepsy related variants[150]. Still, no *in silico* prediction tool is able to capture all potentially pathogenic variants and as CADD scores are only available for missense variants, indels and frameshift variants may be underestimated.

I was not able to Sanger validate and perform inheritance studies on all genetic variants of interest due to lack of DNA (subject 138J with *KIAA2022* variant and subject 00514 with *SYNGAP1* variant), and failed PCR in one case (subject 00526 with *CHD4*). In addition, I did not perform paternity checks on all *de novo* variants.

Many of the genes and variants described in this thesis are novel in their association with MAE and/or with only a handful of previously reported cases. Further functional assessment in a model system of these variants would be the gold standard in understanding whether each

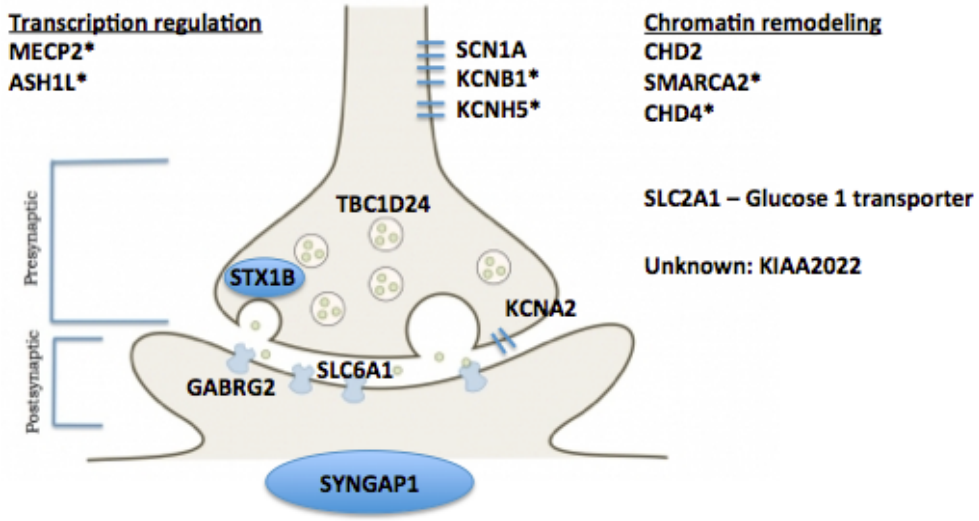
variant cause loss or gain of function. For example, blood RNA is being collected from subject 00525 with a c.1569+7G>T *KCNH5* variant in order to perform expression studies. In addition, given the rarity of these gene variants, web based gene matching platforms (e.g. https://genematcher.org/) can be a useful tool to facilitate collection of larger cohorts.

Last, the limitations described above, plus the rapidly evolving pace in the field of epilepsy genetics, brings home the point of the importance of re-assessment of variants. Hagman *et al*. reported a re-classification rate of 22% of candidate variants following literature review after 12 months[282]. Therefore a re-review of the solved and unsolved cases in this study after a suitable interval would likely lead to fruitful uncovering.

8.4 The genetic heterogeneity of MAE

This study has expanded the genetic heterogeneity of MAE by identifying three further genes novel in association with MAE (*KCNB1, KCNH5*, *MECP2*) and three candidate genes (*SMARCA2*, *ASH1L*, *CHD4*). I have also described a MAE associated phenotype in genes which are known to be associated with congenital syndromes with seizures (*SMARCA2*, *MECP2*). The extent of genetic heterogeneity in MAE is now clear, with each gene accounting for only a small proportion of cases. The major pathways involved, which have also been previously recognised in other epilepsies, are ion channels[283] (*SCN1A*, *GABRG2*, *KCNA2*, *KCNB1*, *KCNH5*), synaptic transmission (*STX1B*, *SYNGAP1*), transcriptional regulation (*MECP2*) and chromatic modification (*CHD2*, *SMARCA2*, *CHD4*, *ASH1L*) (see Figure 8.2).

Figure 8.2 Diagrammatic representation of genes associated with MAE.



*novel MAE associated or candidate genes identified in this study

Ion channelopathies are an established mode of pathogenesis in epilepsy, while chromatin remodelling is more recently recognised. Ion channels are transmembrane proteins that after activation allow the passive flow of ions in and out of cells or cellular organelles following their electrochemical gradients. This flux of ions generates a membrane potential and this can function as neuronal signalling and allow neurotransmitter release. Examples are *SCN1A* and *GABRG2*. *SCN1A* encodes the Nav1.1 channel, one of nine subtypes of voltage gated sodium channels that is preferentially expressed in GABAergic neurons. *GABRG2* encodes the $GABA_A$ receptor, which is the major inhibitory neurotransmitter receptor in the central nervous system. Hence dysfunction of either NaV1.1 channels or $GABA_A$ receptors leads to reduced excitability of GABAergic neurons and increased brain hyperexcitability. As a group, ion channels comprise over 400 members, each with a unique but overlapping functional role. However, they conferred similar variant complexity when unaffected individuals and patients with epilepsy were compared[178]. Therefore, *in silico* modelling of channel variation in cell and network models is important to determine their individual role in pathogenicity, such as multistate structural modelling and voltage clamp analysis of the p.R327H *KCNH5* variant of a patient with epilepsy and autism[231].

Chromatin remodelling is the modification of chromatin architecture from a condensed state by which the DNA wrapped within nucleosomes becomes a transcriptionally accessible state, allowing transcription factors or other DNA binding proteins to access the DNA and control gene expression[277]. Defects in this process can lead to impaired synaptic differentiation and neuronal connectivity in animal models[278]. DNA accessibility is facilitated by chromatin remodelling complexes that utilize ATP to disrupt nucleosome DNA contacts and move nucleosomes[277]. *CHD2*, *SMARCA2* and *CHD4* are all members of the ATPase family, suggesting this pathway as a potential therapeutic target.

The genetic heterogeneity of MAE leads to a phenotypic heterogeneity and reciprocally the reverse is true. Table 8.1 summarises the likely pathogenic or candidate variants identified in this cohort and specific phenotypic features which may offer discriminatory clues to a genetic correlation. In general, a variable epilepsy prognosis was seen (5/10 achieved seizure remission >2 years), although the phenotype leaned towards an EE as is reflected in significant neurodevelopmental comorbidity: ID (11/11), ASD (6/11) and ADHD (3/11), supporting a shared genetic and biological link between these various neurodevelopmental disorders[166]. These neurodevelopmental comorbidities may also be a primary feature of genetic disease rather than secondary to disturbance of brain function due to excessive epileptiform activity. In addition, abnormal neurology was frequently recognised (5/11), and EEG offered atypical

findings for MAE (4/11). Some genes have more distinctive phenotypic features and these are highlighted in italics in Table 8.1.

Table 8.1 Likely pathogenic and candidate variants identified and possible phenotypic clues.

Gene	Subject	Variant identified	MAE course Phenotypic clues	EEG
*ASH1L**	00530	c.G2170T:p.A724S	Unfav MAE ID, ASD, ADHD	GSW and focal spikes
CHD2	00559	c.3734+7A>G	Fav MAE *Atonic myoclonic absence seizures*, ID, ADHD	Polyspike and wave
*CHD4**	00526	c.A2687G:p.H896R	Unfav MAE ID	GSW
KCNB1	00533	c.C916T:p.R306C	Unfav MAE *Infantile onset*, ID	Frequent and brief spikes
KCNH5	00525	c.1569+7G>T	Fav MAE, ID	Generalised and focal spikes
KIAA2022	138J	c.1261_1270del: p.L421fs	Unknown prognosis MAE, ID	GSW and polyspikes
MECP2	291J	c.C673A;p.P225T	Fav MAE ID, *cerebellar signs*	Disorganised background with bilateral GSW
SLC6A1	00595	c.C1155G;p.F385L	Fav MAE Early developmental delay, ID, ASD	GSW provoked by *posterior eye closure sensitivity*
*SMARCA2**	30003301	c.3721C>G;p.Q1241E	Fav MAE *Dysmorphism*, ID, ASD, absent speech, feeding difficulties	GSW and polyspike
SYNGAP1	00512	c.2176_2179del: pR726fs	Unfav MAE *Early developmental delay*, ID, ASD, and *ataxia*	GSW
	00514	c.2562_2578del: p.854_860del	Unfav MAE *Early developmental delay*, ID, ASD, ADHD, *hypotonia and ataxia*	Slow background and GSW

*candidate genes, Fav MAE favourable MAE epilepsy course where seizure remission achieved for > 2 years, Unfav MAE unfavourable MAE epilepsy course with on-going seizures, ID intellectual disability, ASD autism spectrum disorder, ADHD attention deficit hyperactivity disorder, GSW generalised spike wave, distinctive phenotype – genotype correlation are in italics.

8.5 MAE: the bridge between GGE and EE

MAE has phenotypic and genetic overlaps with GGE and EE and can be regarded as the bridge between these two distinct entities, which have a different clinical course, treatment options, and prognostic outcomes. GGE (formerly called the idiopathic generalised epilepsies) represents the most common group of epilepsies accounting for 20 to 30% of all epilepsies[36].

The GGE syndromes encompass a group of epilepsies with a presumed genetic basis and characteristic age related clinical and EEG patterns. There are several sub syndromes within GGE and they are defined as an epilepsy that arise spontaneously without associated structural or neurological signs, or metabolic abnormalities[47]. GGE gives rise to three main seizure types: GTCS, absence seizures and myoclonic jerks, in various combinations depending on syndrome. EEG shows GSW and/or generalised polyspike and wave, and sometimes photosensitivity.

EE is a very different entity and is defined by the ILAE as a process by which the epileptic activity itself may contribute to severe cognitive and behavioural impairments above and beyond what might be expected from the underlying pathology alone and that these might worsen over time[35]. However, in the literature the term EE is also used as a category for severe epilepsy syndromes characterised by multiple seizure types, abundant epileptiform activity, and developmental delay or regression[284], this term is used in this context in the large NGS epilepsy studies. The archetype epilepsy syndromes that are described as EE are LGS and West syndrome, but also MAE.

However MAE can have a clinical course more consistent with a GGE than EE. Cases with MAE typically present in normal developing children and apart from the presence of myoclonic atonic seizures have main seizure types of GTCS, myoclonic seizures and absences seizures similar to GGE. Furthermore, EEG patterns in MAE are similar to GGE. In some cases, seizures are well controlled with minimal comorbidity eventually leading to seizure remission. However in other cases, seizures are multiple and refractory, associated with severe neurodevelopmental comorbidity, EEG shows a slow background, shifting the overall picture to an EE. Table 8.2 compares the main electroclinical features of GGE, using sub syndromes childhood absence epilepsy (CAE) and juvenile myoclonic epilepsy (JME), MAE and LGS as an example of an EE.

Table 8.2 Comparison of electroclinical features of GGE, MAE and LGS.

	GGE[285]	MAE (n=119)	LGS[286]
Age at onset	CAE: 5 – 6 years JME: 10 – 15 years	6 months – 6 years	3 – 5 years
Family history of epilepsy	CAE: 40% JME: 50-60%	40%	Unusual
Development delay prior to seizure onset	No	Rarely	Often
Major seizure types	Myoclonic jerks GTCS, typical absences	Myoclonic atonic, GTCS, Myoclonic jerks, typical and atypical absences	Tonic, atonic, atypical absence
EEG	Normal background Generalised discharges of spike/polyspikes waves	Usually normal background Generalised discharges of spike/polyspikes waves	Slow background Generalised slow spike and wave discharges Fast activity in sleep
Intellectual disability	Rarely	Variable	Most
Seizure remission	CAE: Often JME: Rarely	~ 30%	No

CAE childhood absence epilepsy, JME juvenile myoclonic epilepsy

Along with the phenotypic overlaps, the genetic contribution of both GGE and EE overlaps with that of MAE. The genetics of GGE shows a complex non Mendelian pattern of inheritance from the interactions of several susceptibility genes of minor effect along with environmental factors[287]. The high proportion of relatives with epilepsy in GGE and MAE supports this hypothesis. This complex genetic architecture remains difficult to unravel. Linkage analysis has identified several susceptibility loci for GGE sub syndromes[102,288,289], for example EF hand domain containing protein (*EFHC1*) at 6p12-p11 locus[290] and bromodomain containing protein (*BRD2*) at the 6p21 loci for JME[289], but they are poorly replicated due to the high degree of genetic complexity and heterogeneity. Large scale association analysis did not identify specific rare variants to confer significant risk when using a genome wide approach[291], or through a subset of ion channel genes[178]. Genome wide association analysis of 3020 GGE cases implicated susceptibility loci at 2p16.1 and 17q21.32[103], and at 2q24.3 with a larger, more heterogenous overlapping cohort of 8696 cases[292]. Recently, a case control sequencing study demonstrated that patients with GGE have ultra rare genetic variants (MAF<0.001) in known epilepsy genes relative to the rest of protein coding genes (P=5.8 ×10^{-8})[283], suggesting a connection between GGE and rare epilepsies. A monogenic aetiology per se has been identified in a small number of rare multiplex families. Of note, Wallace *et al.* identified *GABRG2* mutations in a large epilepsy pedigree with CAE and febrile seizures that had a single member with MAE[9]. The co-occurrence of GGE and MAE within the same family was also evident in this study in four cases (subjects S1040, 00518, 00574 and 00602), suggesting that

similar genetic influences can account for phenotypic variability consisting of MAE and GGE. Furthermore, *SLC2A1* pathogenic variants have been identified in patients with GGE and MAE[13,267]. Evidence of a shared genetic aetiology is also demonstrated with EE.

All reported genes associated with MAE (see Table 1.8), as well as the novel MAE epilepsy gene associations identified in this study; *KCNB1*, *MECP2* and *KCNH5*, can result in an EE-like phenotype when mutated. *De novo* dominant mutations are the most frequently identified genetic aetiology in EE. Some genes, e.g. *SYNGAP1* and *CHD2* were first identified through NGS studies using a *de novo* trio sequencing and/or a candidate gene approach of large cohorts of EE patients[12,101]. Candidate genes were sometimes selected through CNVs. For example, target sequencing of *SLC6A1* was performed as it was one of two genes in a *de novo* 3p25.3 CNV deletion in a single case with MAE, which subsequently led to its discovery of association in a further 6/160 MAE patients[17]. This is in contrast to GGE where recurrent deletions of 15q13.3, 15q11.2 and 16p13.11 have been identified as specific risk factors[91,92]. In EE, these CNV hotspots are rarely seen and it is the overlapping genes within CNVs that offer a pathogenic role. In MAE, both these situations have been reported: a *de novo* 16p13.11 GGE hotspot deletion was identified in a patient with MAE[95], and CNVs overlapping known EE genes of *STX1B* and *TBC1D24* have been reported[21,22].

The genetic landscape of MAE can be conceived as a continuum starting at one end with a fully penetrant monogenic epilepsy (like EE) progressing through to those caused by less penetrant alleles of large effect and then through oligogenic alleles with smaller effect (presumably like GGE). The overall MAE phenotype of the individual depends on the unique combination of genes and environmental factors involved, and can extend from an EE-like, to a GGE-like disease course. Indeed, this is congruous with Herman Doose's original account of MAE as a paradigmatic example of a multifactorial determined disease with uniting features[28].

8.6 Applications and future directions

This work demonstrates the phenotypic and genetic spectrum of MAE, both of which will continue to grow through continued progress in genomic research. The overall yield of identifying a genetic aetiology in this MAE cohort was 17.4% (19/109). This is comparable to recent epilepsy panel gene studies in patients with heterogenous phenotypes of severe developmental delay, familial epilepsies and EE of 18% by Trump *et al.*[293] and 23% by Møller *et al*.[294]. A gene panel is a frequently used NGS technique where only genes of interest are interrogated for pathogenic variants, and have replaced single gene testing in many centres.

The candidate genes proposed in this project, *SMARCA2*, *ASH1L* and *CHD4* could be added to an epilepsy panel to help identify more patients.

Another approach is to apply exome sequencing used here, as a dual research and clinical tool where a patient's exome can be analysed for known genetic variants as well as uncover novel candidate genes. In addition, exomes can be continually re-analysed for newly identified discoveries until a genetic aetiology is identified. Pipelines for the pathogenic validation of variants have been developed with specific scoring criteria[250,282,295]. Notable, is the American College of Medical Genetics and Genomics guidelines for interpretation of sequence variant, which is frequently cited for clinical use[295]. However, novel ethical challenges are also presented through this move of genomic information to the clinical domain such as translating incidental findings, and translating variants of uncertain significance to patients.

Even without direct therapeutic impact, a genetic diagnosis can significantly impact on families. For example, a *de novo* genetic diagnosis can help relieve guilt from false attribution form the parents' consciousness that an early life event could have produced symptoms, and offer informed guidance for parents intending to extend their family. In addition, ending the diagnostic odyssey avoids further investigations for an underlying aetiology, which is both costly and distressing for the child and family.

The work continues also for the remaining 83% of unsolved MAE cases. Several lines of enquiry remain. The impact of variants in non-coding regions, regulatory regions and microRNAs remains to be explored, and may be revealed with increased whole genome sequencing. MicroRNAs are small non-coding RNAs that control the expression levels of multiple proteins by decreasing mRNA stability and translation. Manipulation of 12 microRNAs in mouse models were showed to have effects on EEG, seizures and histopathology[296,297]. Patients with temporal lobe epilepsy were found to have increased micro RNA-181a levels in the hippocampus[296]. A further example was by Panjwani *et al.*, who demonstrated that the homozygous T allele at microRNA-328 binding site conferred increased risk to EEG features of centro temporal spikes in patients with rolandic epilepsy (P=2.6 x 10^{-4})[298]. The specific cell types and targets for specific microRNAs remains to be identified and offers an untapped research resource in contributing to the pathogenesis of epilepsy.

Further genetic studies exploring other models will shed more light. CNVs in my cohort remains to be thoroughly investigated and were not explored as first line due to the limitations in calling CNVs with exome sequencing data and time constraints. The Epilepsy

Phenome/Genome Project and Epi4K consortium demonstrated the level of false positives with this method and were able to validate by array based methods only 66% (53/80) of inherited CNVs and 24% (5/21) of *de novo* CNVs called from exome sequencing data in patients with EE[299]. The ascertainment of rare multiplex MAE families may offer further insights into allelic contribution of specific phenotypic and EEG features, for example transmission of specific EEG features or power spectra frequency bands, and aid in offering causal variant discovery in alleles with incomplete penetrance.

Somatic mosaic variants in epilepsy are not well studied. Somatic mosaicism occurs in embryogenesis, resulting in variants confined to a proportion of all cell types, specific body regions or tissue lineages. Different tissues can show differing levels of mosaicism such that low levels of an abnormal allele are difficult to detect. Somatic mosaic variants are similar to *de novo* variants in that they have not been subject to purifying selection and consequently are more likely to be deleterious than inherited variants. Somatic variants e.g. *MTOR*, *PIK3R2* have been associated with brain malformations and epilepsy, their identification was aided by the availability of testing brain tissue following surgery in these patients, which would otherwise have been difficult to assess. However, Halvorsen *et al.* constructed a mosaic transmission screen using trio exome sequencing data on nine cases with sudden unexplained death in childhood and 338 cases with EE. In brief, the screen extracted novel nonsynonymous variant calls from the proband and parents, and then performed a binomial exact test on the mutant allele read ratio for the carrier parent and proband child to determine the probability of obtaining the observed read ratio given an expected value of 0.5 for heterozygous variants. They identified six parental-mosaic transmissions across the two cohorts (estimated 0.02 transmission per trio)[83]. This mechanism appears lower than *de novo* variants but remains applicable in epilepsy.

Collaborative approaches will be necessary to overcome the difficulties due to genetic heterogeneity and rarity of MAE, in order to make fundamental leaps in understanding. A proposed model is a 'lumping and splitting' experimental design where larger cohorts are assembled through multicentre collaborations in order to generate sufficient power for gene identification followed by fractionation into specific gene/allele related phenotypic characteristics[165,292]. The Epilepsy Genetics Initiative (www.cureepilepsy.org/egi) seeks to grant access to exome databases to enable collaborations and collection of gene cohorts. The collection of patients with specific rare pathogenic gene variants can also be achieved through patient-gene registries (www.geneticepilepsy.com) and gene matching web platforms (http://genematcher.org). The accumulation of larger well characterised cohorts can lead to

delineation of the natural history of specific gene related epilepsies, allow more precise phenotype quantification, longitudinal studies on specific comorbidities, and investigate phenotype genotype correlation pertaining to position dependent effects or variant type effects. Additionally, it can lead to more comprehensive investigations in understanding of disease mechanisms, development of novel treatments and clinical trials; adding to the implementation of precision medicine.

Precision medicine is already a reality in epilepsy genetics. For example, the ketogenic diet is used to treat *SLC2A1* associated epilepsy and choice of antiepileptic drugs is impacted through recognition of specific epilepsy associated genes (S*CN1A*, *SCN2A, SCN8A, KCNQ2*). This targeted therapy can be taken on by different approaches: (1) use of pharmacological agents that target the abnormal protein or disrupted pathway, (2) use of induced pluripotent stem cells for in vitro disease modelling, tagged with reprogramming technologies in genetic modification e.g. CRISPR and (3) gene therapy involving modification of transcription of a mutant gene, or delivery or expression of a wild type gene[79]. The identification of chromatin remodelling genes in this project highlights this pathway as a possible therapeutic target.

Through the synthesis of these different approaches, there is much promise in the prospect of unravelling the phenotypic and genetic spectrum of MAE even further, and extending these findings to benefit patients.

References

1. Doose H, Baier WK. Epilepsy with primarily generalized myoclonic-astatic seizures: a genetically determined disease. Eur J Pediatr. 1987;146(6):550-4.
2. Kaminska A, Ickowicz A, Plouin P, Bru MF, Dellatolas G, Dulac O. Delineation of cryptogenic Lennox-Gastaut syndrome and myoclonic astatic epilepsy using multiple correspondence analysis. Epilepsy Res. 1999;36(1):15-29.
3. Guerrini R, Aicardi J. Epileptic encephalopathies with myoclonic seizures in infants and children (severe myoclonic epilepsy and myoclonic-astatic epilepsy). J Clin Neurophysiol. 2003;20(6):449-61.
4. Oguni H, Fukuyama Y, Tanaka T, Hayashi K, Funatsuka M, Sakauchi M, et al. Myoclonic-astatic epilepsy of early childhood--clinical and EEG analysis of myoclonic-astatic seizures, and discussions on the nosology of the syndrome. Brain Dev. 2001;23(7):757-64.
5. Doose H, Gerken H, Leonhardt R, Volzke E, Volz C. Centrencephalic myoclonic-astatic petit mal. Clinical and genetic investigation. Neuropadiatrie. 1970;2(1):59-78.
6. Berkovic SF, Howell RA, Hay DA, Hopper JL. Epilepsies in twins: genetics of the major epilepsy syndromes. Ann Neurol. 1998;43(4):435-45.
7. Kjeldsen MJ, Corey LA, Christensen K, Friis ML. Epileptic seizures and syndromes in twins: the importance of genetic factors. Epilepsy Res. 2003;55(1-2):137-46.
8. Wallace RH, Wang DW, Singh R, Scheffer IE, George AL, Jr., Phillips HA, et al. Febrile seizures and generalized epilepsy associated with a mutation in the Na+-channel beta1 subunit gene SCN1B. Nat Genet. 1998;19(4):366-70.
9. Wallace RH, Marini C, Petrou S, Harkin LA, Bowser DN, Panchal RG, et al. Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet. 2001;28(1):49-52.
10. Escayg A, Heils A, MacDonald BT, Haug K, Sander T, Meisler MH. A novel SCN1A mutation associated with generalized epilepsy with febrile seizures plus--and prevalence of variants in patients with epilepsy. Am J Hum Genet. 2001;68(4):866-73. Epub 2001 Mar 14.
11. Ohmori I, Ouchida M, Ohtsuka Y, Oka E, Shimizu K. Significant correlation of the SCN1A mutations and severe myoclonic epilepsy in infancy. Biochem Biophys Res Commun. 2002;295(1):17-23.
12. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013;45(7):825-30.
13. Mullen SA, Marini C, Suls A, Mei D, Della Giustina E, Buti D, et al. Glucose transporter 1 deficiency as a treatable cause of myoclonic astatic epilepsy. Arch Neurol. 2011;68(9):1152-5. doi: 10.001/archneurol.2011.102. Epub May 9.
14. Trivisano M, Striano P, Sartorelli J, Giordano L, Traverso M, Accorsi P, et al. CHD2 mutations are a rare cause of generalized epilepsy with myoclonic-atonic seizures. Epilepsy Behav. 2015;51:53-6.(doi):10.1016/j.yebeh.2015.06.029. Epub Aug 7.
15. Thomas RH, Zhang LM, Carvill GL, Archer JS, Heavin SB, Mandelstam SA, et al. CHD2 myoclonic encephalopathy is frequently associated with self-induced seizures. Neurology. 2015;84(9):951-8. doi: 10.1212/WNL.0000000000001305. Epub 2015 Feb 11.
16. Mignot C, von Stulpnagel C, Nava C, Ville D, Sanlaville D, Lesca G, et al. Genetic and neurodevelopmental spectrum of SYNGAP1-associated intellectual disability and epilepsy. J Med Genet. 2016;53(8):511-22. doi: 10.1136/jmedgenet-2015-103451. Epub 2016 Mar 17.
17. Carvill GL, McMahon JM, Schneider A, Zemel M, Myers CT, Saykally J, et al. Mutations in the GABA Transporter SLC6A1 Cause Epilepsy with Myoclonic-Atonic Seizures. Am J Hum Genet. 2015;8(15):00069-5.
18. Larsen J, Johannesen KM, Ek J, Tang S, Marini C, Blichfeldt S, et al. The role of SLC2A1 mutations in myoclonic astatic epilepsy and absence epilepsy, and the estimated frequency of GLUT1 deficiency syndrome. Epilepsia. 2015;56(12):e203-8. doi: 10.1111/epi.13222. Epub 2015 Nov 5.

19. Syrbe S, Hedrich UB, Riesch E, Djemie T, Muller S, Moller RS, et al. De novo loss- or gain-of-function mutations in KCNA2 cause epileptic encephalopathy. Nat Genet. 2015;47(4):393-9. doi: 10.1038/ng.3239. Epub 2015 Mar 9.
20. Schubert J, Siekierska A, Langlois M, May P, Huneau C, Becker F, et al. Mutations in STX1B, encoding a presynaptic protein, cause fever-associated epilepsy syndromes. Nat Genet. 2014;46(12):1327-32. doi: 10.038/ng.3130. Epub 2014 Nov 2.
21. Vlaskamp DR, Rump P, Callenbach PM, Vos YJ, Sikkema-Raddatz B, van Ravenswaaij-Arts CM, et al. Haploinsufficiency of the STX1B gene is associated with myoclonic astatic epilepsy. Eur J Paediatr Neurol. 2016;8(16):00003-9.
22. Balestrini S, Milh M, Castiglioni C, Luthy K, Finelli MJ, Verstreken P, et al. TBC1D24 genotype-phenotype correlation: Epilepsies and other neurologic features. Neurology. 2016;87(1):77-85. doi: 10.1212/WNL.0000000000002807. Epub 2016 Jun 8.
23. de Lange IM, Helbig KL, Weckhuysen S, Moller RS, Velinov M, Dolzhanskaya N, et al. De novo mutations of KIAA2022 in females cause intellectual disability and intractable epilepsy. J Med Genet. 2016;29(103909):2016-103909.
24. Lennox WG, Davis JP. Clinical correlates of the fast and the slow spike-wave electroencephalogram. Pediatrics. 1950;5(4):626-44.
25. Gastraut H, Roger J, Soulayrol R, Tassinari CA, Regis H, Dravet C, et al. Childhood epileptic encephalopathy with diffuse slow spike-waves (otherwise known as "petit mal variant") or Lennox syndrome. Epilepsia. 1966;7(2):139-79.
26. Harper JR. True myoclonic epilepsy in childhood. Arch Dis Child. 1968;43(227):28-35.
27. Kruse R. [The myoclonic astatic petit mal. Clinical course of small epileptic seizures in childhood. With an introduction by Prof. Dr. Dietrich Janz]. Monogr Gesamtgeb Neurol Psychiatr. 1968;124:1-126.
28. Roger J BM, Dravet CH, Dreifuss FE, Perret A, Wolf P. Epileptic syndromes in infancy, childhood and adolescence (2nd edition). John Libbey and Company Ltd. 1992.
29. Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy. Epilepsia. 1989;30(4):389-99.
30. Oguni H, Tanaka T, Hayashi K, Funatsuka M, Sakauchi M, Shirakawa S, et al. Treatment and long-term prognosis of myoclonic-astatic epilepsy of early childhood. Neuropediatrics. 2002;33(3):122-32.
31. Nabbout R, Kozlovski A, Gennaro E, Bahi-Buisson N, Zara F, Chiron C, et al. Absence of mutations in major GEFS+ genes in myoclonic astatic epilepsy. Epilepsy Res. 2003;56(2-3):127-33.
32. Kilaru S, Bergqvist AG. Current treatment of myoclonic astatic epilepsy: clinical experience at the Children's Hospital of Philadelphia. Epilepsia. 2007;48(9):1703-7.
33. Trivisano M, Specchio N, Cappelletti S, Di Ciommo V, Claps D, Specchio LM, et al. Myoclonic astatic epilepsy: an age-dependent epileptic syndrome with favorable seizure outcome but variable cognitive evolution. Epilepsy Res. 2011;97(1-2):133-41.
34. Caraballo RH, Chamorro N, Darra F, Fortini S, Arroyo H. Epilepsy with myoclonic atonic seizures: an electroclinical study of 69 patients. Pediatr Neurol. 2013;48(5):355-62.
35. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. Epilepsia. 2010;51(4):676-85. doi: 10.1111/j.528-67.2010.02522.x. Epub 2010 Feb 26.
36. Jallon P, Loiseau P, Loiseau J. Newly diagnosed unprovoked epileptic seizures: presentation at diagnosis in CAROLE study. Coordination Active du Reseau Observatoire Longitudinal de l' Epilepsie. Epilepsia. 2001;42(4):464-75.
37. Doose H, Sitepu B. Childhood epilepsy in a German city. Neuropediatrics. 1983;14(4):220-4.

38. Moeller F, Groening K, Moehring J, Muhle H, Wolff S, Jansen O, et al. EEG-fMRI in myoclonic astatic epilepsy (Doose syndrome). Neurology. 2014;82(17):1508-13. doi: 10.212/WNL.0000000000000359. Epub 2014 Apr 2.
39. Aghakhani Y, Bagshaw AP, Benar CG, Hawco C, Andermann F, Dubeau F, et al. fMRI activation during spike and wave discharges in idiopathic generalized epilepsy. Brain. 2004;127(Pt 5):1127-44. Epub 2004 Mar 19.
40. Blume WT, Luders HO, Mizrahi E, Tassinari C, van Emde Boas W, Engel J, Jr. Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology. Epilepsia. 2001;42(9):1212-8.
41. Hirano Y, Oguni H, Funatsuka M, Imai K, Osawa M. Differentiation of myoclonic seizures in epileptic syndromes: a video-polygraphic study of 26 patients. Epilepsia. 2009;50(6):1525-35.
42. Arzimanoglou A GR, Aicardi J. Aicardi's epilepsy in children. Lippincott Williams and Wilkins, Philadelphia, PA. 2004.
43. Bonanni P, Parmeggiani L, Guerrini R. Different neurophysiologic patterns of myoclonus characterize Lennox-Gastaut syndrome and myoclonic astatic epilepsy. Epilepsia. 2002;43(6):609-15.
44. Dulac O, N'Guyen T. The Lennox-Gastaut syndrome. Epilepsia. 1993;34(Suppl 7):S7-17.
45. Trinka E, Cock H, Hesdorffer D, Rossetti AO, Scheffer IE, Shinnar S, et al. A definition and classification of status epilepticus--Report of the ILAE Task Force on Classification of Status Epilepticus. Epilepsia. 2015;56(10):1515-23. doi: 10.111/epi.13121. Epub 2015 Sep 4.
46. Dragoumi P, Chivers F, Brady M, Craft S, Mushati D, Venkatachalam G, et al. Epilepsy with myoclonic-atonic seizures (Doose syndrome): When video-EEG polygraphy holds the key to syndrome diagnosis. Epilepsy Behav Case Rep. 2016;5:31-3.(doi):10.1016/j.ebcr.2015.10.001. eCollection 6.
47. Engel J, Jr. Report of the ILAE classification core group. Epilepsia. 2006;47(9):1558-68.
48. Inoue T, Ihara Y, Tomonoh Y, Nakamura N, Ninomiya S, Fujita T, et al. Early onset and focal spike discharges as indicators of poor prognosis for myoclonic-astatic epilepsy. Brain Dev. 2014;36(7):613-9. doi: 10.1016/j.braindev.2013.08.009. Epub Sep 19.
49. Stephani U. The natural history of myoclonic astatic epilepsy (Doose syndrome) and Lennox-Gastaut syndrome. Epilepsia. 2006;47(Suppl 2):53-5.
50. Nickels KC, Zaccariello MJ, Hamiwka LD, Wirrell EC. Cognitive and neurodevelopmental comorbidities in paediatric epilepsy. Nat Rev Neurol. 2016;12(8):465-76. doi: 10.1038/nrneurol.2016.98. Epub Jul 22.
51. Nolte R, Wolff M. Behavioural and developmental aspects of primary generalized myoclonic-astatic epilepsy. Epilepsy Res Suppl. 1992;6:175-83.
52. Tang S, Hughes E, Lascelles K, Simpson MA, Pal DK. New SMARCA2 mutation in a patient with Nicolaides-Baraitser syndrome and myoclonic astatic epilepsy. Am J Med Genet A. 2016;26(10):37935.
53. Filippini M, Boni A, Dazzani G, Guerra A, Gobbi G. Neuropsychological findings: myoclonic astatic epilepsy (MAE) and Lennox-Gastaut syndrome (LGS). Epilepsia. 2006;47(Suppl 2):56-9.
54. van Rijckevorsel K. Cognitive problems related to epilepsy syndromes, especially malignant epilepsies. Seizure. 2006;15(4):227-34. Epub 2006 Mar 24.
55. Smith DB, Craft BR, Collins J, Mattson RH, Cramer JA. Behavioral characteristics of epilepsy patients compared with normal controls. Epilepsia. 1986;27(6):760-8.
56. Vingerhoets G. Cognitive effects of seizures. Seizure. 2006;15(4):221-6. Epub 2006 Mar 20.
57. Kelley SA, Kossoff EH. Doose syndrome (myoclonic-astatic epilepsy): 40 years of progress. Dev Med Child Neurol. 2010;52(11):988-93.
58. Ohtsuka Y, Yoshinaga H, Kobayashi K, Ogino T, Oka M, Ito M. Diagnostic issues and treatment of cryptogenic or symptomatic generalized epilepsies. Epilepsy Res. 2006;70(Suppl 1):S132-40. Epub 2006 Jul 11.

59. Cvetkovska E, Panov S. Possible genetic anticipation in families with idiopathic generalised epilepsy. Epileptic Disord. 2011;13(2):150-4. doi: 10.1684/epd.2011.0441.
60. Ottman R, Barker-Cummings C, Leibson CL, Vasoli VM, Hauser WA, Buchhalter JR. Accuracy of family history information on epilepsy and other seizure disorders. Neurology. 2011;76(4):390-6. doi: 10.1212/WNL.0b013e3182088286.
61. Berkovic SF, Mulley JC, Scheffer IE, Petrou S. Human epilepsies: interaction of genetic and acquired factors. Trends Neurosci. 2006;29(7):391-7.
62. Durner M, Keddache MA, Tomasini L, Shinnar S, Resor SR, Cohen J, et al. Genome scan of idiopathic generalized epilepsy: evidence for major susceptibility gene and modifying genes influencing the seizure type. Ann Neurol. 2001;49(3):328-35.
63. Pal DK, Durner M, Klotz I, Dicker E, Shinnar S, Resor S, et al. Complex inheritance and parent-of-origin effect in juvenile myoclonic epilepsy. Brain Dev. 2006;28(2):92-8. Epub 2006 Jan 18.
64. Gottesman, II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry. 2003;160(4):636-45.
65. Waltz S, Christen HJ, Doose H. The different patterns of the photoparoxysmal response--a genetic study. Electroencephalogr Clin Neurophysiol. 1992;83(2):138-45.
66. Pinto D, Westland B, de Haan GJ, Rudolf G, da Silva BM, Hirsch E, et al. Genome-wide linkage scan of epilepsy-related photoparoxysmal electroencephalographic response: evidence for linkage on chromosomes 7q32 and 16p13. Hum Mol Genet. 2005;14(1):171-8. Epub 2004 Nov 17.
67. Tauer U, Lorenz S, Lenzen KP, Heils A, Muhle H, Gresch M, et al. Genetic dissection of photosensitivity and its relation to idiopathic generalized epilepsy. Ann Neurol. 2005;57(6):866-73.
68. Doose H, Baier WK. Theta rhythms in the EEG: a genetic trait in childhood epilepsy. Brain Dev. 1988;10(6):347-54.
69. Helbig I, Tayoun AA. Understanding Genotypes and Phenotypes in Epileptic Encephalopathies. Mol Syndromol. 2016;7(4):172-81. Epub 2016 Aug 20.
70. Dimova PS, Yordanova I, Bojinova V, Jordanova A, Kremenski I. Generalized epilepsy with febrile seizures plus: novel SCN1A mutation. Pediatr Neurol. 2010;42(2):137-40. doi: 10.1016/j.pediatrneurol.2009.09.007.
71. Ebach K, Joos H, Doose H, Stephani U, Kurlemann G, Fiedler B, et al. SCN1A mutation analysis in myoclonic astatic epilepsy and severe idiopathic generalized epilepsy of infancy with generalized tonic-clonic seizures. Neuropediatrics. 2005;36(3):210-3.
72. Yordanova I, Todorov T, Dimova P, Hristova D, Tincheva R, Litvinenko I, et al. One novel Dravet syndrome causing mutation and one recurrent MAE causing mutation in SCN1A gene. Neurosci Lett. 2011;494(2):180-3. doi: 10.1016/j.neulet.2011.03.008. Epub Mar 15.
73. Wang D, Pascual JM, De Vivo D. Glucose Transporter Type 1 Deficiency Syndrome. 1993.
74. Shen D, Hernandez CC, Shen W, Hu N, Poduri A, Shiedley B, et al. De novo GABRG2 mutations associated with epileptic encephalopathies. Brain. 2016;17.
75. Palmer S, Towne MC, Pearl PL, Pelletier RC, Genetti CA, Shi J, et al. SLC6A1 Mutation and Ketogenic Diet in Epilepsy With Myoclonic-Atonic Seizures. Pediatr Neurol. 2016;28(16):30072-8.
76. Epi KC, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, et al. De novo mutations in epileptic encephalopathies. Nature. 2013.
77. Suls A, Jaehn JA, Kecskes A, Weber Y, Weckhuysen S, Craiu DC, et al. De novo loss-of-function mutations in CHD2 cause a fever-sensitive myoclonic epileptic encephalopathy sharing features with Dravet syndrome. Am J Hum Genet. 2013;93(5):967-75. doi: 10.1016/j.ajhg.2013.09.017. Epub Oct 24.
78. Galizia EC, Myers CT, Leu C, de Kovel CG, Afrikanova T, Cordero-Maldonado ML, et al. CHD2 variants are a risk factor for photosensitivity in epilepsy. Brain. 2015;138(Pt 5):1198-207. doi: 10.093/brain/awv052. Epub 2015 Mar 17.

79. Pinto AM, Bianciardi L, Mencarelli MA, Imperatore V, Di Marco C, Furini S, et al. Exome sequencing analysis in a pair of monozygotic twins re-evaluates the genetics behind their intellectual disability and reveals a CHD2 mutation. Brain Dev. 2016;38(6):590-6. doi: 10.1016/j.braindev.2015.12.006. Epub 6 Jan 2.
80. Lemke JR, Riesch E, Scheurenbrand T, Schubach M, Wilhelm C, Steiner I, et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia. 2012;53(8):1387-98. doi: 10.111/j.528-167.2012.03516.x. Epub 2012 May 21.
81. Corbett MA, Bellows ST, Li M, Carroll R, Micallef S, Carvill GL, et al. Dominant KCNA2 mutation causes episodic ataxia and pharmacoresponsive epilepsy. Neurology. 2016;87(19):1975-84. Epub 2016 Oct 12.
82. Vlaskamp DR, Rump P, Callenbach PM, Vos YJ, Sikkema-Raddatz B, van Ravenswaaij-Arts CM, et al. Haploinsufficiency of the STX1B gene is associated with myoclonic astatic epilepsy. Eur J Paediatr Neurol. 2016;20(3):489-92. doi: 10.1016/j.ejpn.2015.12.014. Epub 6 Jan 8.
83. Halvorsen M, Petrovski S, Shellhaas R, Tang Y, Crandall L, Goldstein D, et al. Mosaic mutations in early-onset genetic diseases. Genet Med. 2016;18(7):746-9. doi: 10.1038/gim.2015.155. Epub Dec 30.
84. Corbett MA, Bahlo M, Jolly L, Afawi Z, Gardner AE, Oliver KL, et al. A focal epilepsy and intellectual disability syndrome is due to a mutation in TBC1D24. Am J Hum Genet. 2010;87(3):371-5. doi: 10.1016/j.ajhg.2010.08.001.
85. Cantagrel V, Lossi AM, Boulanger S, Depetris D, Mattei MG, Gecz J, et al. Disruption of a new X linked gene highly expressed in brain in a family with two mentally retarded males. J Med Genet. 2004;41(10):736-42.
86. Farach LS, Northrup H. KIAA2022 nonsense mutation in a symptomatic female. Am J Med Genet A. 2016;170(3):703-6. doi: 10.1002/ajmg.a.37479. Epub 2015 Nov 17.
87. Kang JQ, Macdonald RL. Molecular Pathogenic Basis for GABRG2 Mutations Associated With a Spectrum of Epilepsy Syndromes, From Generalized Absence Epilepsy to Dravet Syndrome. JAMA Neurol. 2016;73(8):1009-16. doi: 10.1/jamaneurol.2016.0449.
88. Reinthaler EM, Dejanovic B, Lal D, Semtner M, Merkler Y, Reinhold A, et al. Rare variants in gamma-aminobutyric acid type A receptor genes in rolandic epilepsy and related syndromes. Ann Neurol. 2015;77(6):972-86. doi: 10.1002/ana.24395. Epub 2015 Mar 28.
89. Boillot M, Morin-Brureau M, Picard F, Weckhuysen S, Lambrecq V, Minetti C, et al. Novel GABRG2 mutations cause familial febrile seizures. Neurol Genet. 2015;1(4):e35. doi: 10.1212/NXG.0000000000000035. eCollection 2015 Dec.
90. Mefford HC, Yendle SC, Hsu C, Cook J, Geraghty E, McMahon JM, et al. Rare copy number variants are an important cause of epileptic encephalopathies. Ann Neurol. 2011;70(6):974-85.
91. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet. 2009;41(2):160-2. doi: 10.1038/ng.292. Epub 2009 Jan 11.
92. de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain. 2010;133(Pt 1):23-32. doi: 10.1093/brain/awp262. Epub 2009 Oct 20.
93. Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. 2010;6(5):e1000962. doi: 10.1371/journal.pgen..
94. Striano P, Coppola A, Paravidino R, Malacarne M, Gimelli S, Robbiano A, et al. Clinical significance of rare copy number variations in epilepsy: a case-control survey using microarray-based comparative genomic hybridization. Arch Neurol. 2012;69(3):322-30. doi: 10.1001/archneurol.2011.1999. Epub 2011 Nov 14.
95. Helbig I, Swinkels ME, Aten E, Caliebe A, van 't Slot R, Boor R, et al. Structural genomic variation in childhood epilepsies with complex phenotypes. Eur J Hum Genet. 2014;22(7):896-901. doi: 10.1038/ejhg.2013.262. Epub Nov 27.

96. Ottaviani V, Bartocci A, Pantaleo M, Giglio S, Cecconi M, Verrotti A, et al. MYOCLONIC ASTATIC EPILEPSY IN A PATIENT WITH A DE NOVO 4q21.22q21.23 MICRODUPLICATION. Genet Couns. 2015;26(3):327-32.
97. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010;42(9):790-3. doi: 10.1038/ng.646. Epub 2010 Aug 15.
98. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012;20(5):490-7.
99. Heron SE, Smith KR, Bahlo M, Nobili L, Kahana E, Licchetta L, et al. Missense mutations in the sodium-gated potassium channel gene KCNT1 cause severe autosomal dominant nocturnal frontal lobe epilepsy. Nat Genet. 2012;44(11):1188-90.
100. Berger I, Dor T, Halvardson J, Edvardson S, Shaag A, Feuk L, et al. Intractable epilepsy of infancy due to homozygous mutation in the EFHC1 gene. Epilepsia. 2012;53(8):1436-40.
101. Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501(7466):217-21. doi: 10.1038/nature12439. Epub 2013 Aug 11.
102. Leu C, de Kovel CG, Zara F, Striano P, Pezzella M, Robbiano A, et al. Genome-wide linkage meta-analysis identifies susceptibility loci at 2q34 and 13q31.3 for genetic generalized epilepsies. Epilepsia. 2012;53(2):308-18. doi: 10.1111/j.528-67.2011.03379.x. Epub 2012 Jan 13.
103. Steffens M, Leu C, Ruppert AK, Zara F, Striano P, Robbiano A, et al. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Hum Mol Genet. 2012;21(24):5359-72. doi: 10.1093/hmg/dds373. Epub 2012 Sep 4.
104. Davidson AJ, Disma N, de Graaff JC, Withington DE, Dorris L, Bell G, et al. Neurodevelopmental outcome at 2 years of age after general anaesthesia and awake-regional anaesthesia in infancy (GAS): an international multicentre, randomised controlled trial. Lancet. 2016;387(10015):239-50. doi: 10.1016/S0140-6736(15)00608-X. Epub 2015 Nov 4.
105. N B. Bayley scales of infant and toddler development. 3rd ed. San Antonio: Harcourt Assessment Inc. 2006.
106. Berg AT, Plioplys S, Tuchman R. Risk and correlates of autism spectrum disorder in children with epilepsy: a community-based study. J Child Neurol. 2011;26(5):540-7. doi: 10.1177/0883073810384869. Epub 2011 Mar 18.
107. Skuse DH, Mandy WP, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. Br J Psychiatry. 2005;187:568-72.
108. Association AP. Diagnostic and statisical manual of mental disorders 4th ed Washington DC. 1994.
109. Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2004;43(5):548-58.
110. Slappendel G, Mandy W, van der Ende J, Verhulst FC, van der Sijde A, Duvekot J, et al. Utility of the 3Di Short Version for the Diagnostic Assessment of Autism Spectrum Disorder and Compatibility with DSM-5. J Autism Dev Disord. 2016;46(5):1834-46. doi: 10.007/s10803-016-2713-9.
111. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. Int Rev Psychiatry. 2003;15(1-2):166-72.
112. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38(5):581-6.
113. CK C. Conner's Rating Scales Manual. Instruments for use with children and adolescents, New York: Mutli-Health Systems, Inc. 1989.
114. Harrison P OT. Adaptive Behaviour Assessment System (2nd Edition) manual. USA: Western Psycholoigcal Services. 2003.

115. Abend NS, Wagenman KL, Blake TP, Schultheis MT, Radcliffe J, Berg RA, et al. Electrographic status epilepticus and neurobehavioral outcomes in critically ill children. Epilepsy Behav. 2015;49:238-44.(doi):10.1016/j.yebeh.2015.03.013. Epub Apr 20.
116. Skuse DH, Mandy W, Steer C, Miller LL, Goodman R, Lawrence K, et al. Social communication competence and functional adaptation in a general population of children: preliminary evidence for sex-by-verbal IQ differential risk. J Am Acad Child Adolesc Psychiatry. 2009;48(2):128-37. doi: 10.1097/CHI.0b013e31819176b8.
117. Chuthapisith J, Taycharpipranai P, Ruangdaraganon N, Warrington R, Skuse D. Translation and validation of the developmental, dimensional and diagnostic interview (3Di) for diagnosis of autism spectrum disorder in Thai children. Autism. 2012;16(4):350-6. doi: 10.1177/1362361311433770. Epub 2012 Mar 7.
118. Meltzer HG, R.; Goodman, R.; Ford, F. Mental health of children and adolescents in Great Britain. London: The Stationery Office. 2000.
119. Iber C A-IS, Chesson A L, Quan S F. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specification, 1st edn. American Academy of Sleep Medicine, Westchester, IL. 2007.
120. Gibbs EL GG. Diagnostic and localizing value of electroencephalographic studies in sleep. J Nerv Ment Dis. 1947;26:366-76.
121. Bazil CW, Walczak TS. Effects of sleep and sleep stage on epileptic and nonepileptic seizures. Epilepsia. 1997;38(1):56-62.
122. Degen R, Degen HE. Sleep and sleep deprivation in epileptology. Epilepsy Res Suppl. 1991;2:235-60.
123. Buchthal F, Lennox M. The EEG effect of metrazol and photic stimulation in 682 normal subjects. Electroencephalogr Clin Neurophysiol. 1953;5(4):545-58.
124. Gregory RP, Oates T, Merry RT. Electroencephalogram epileptiform abnormalities in candidates for aircrew training. Electroencephalogr Clin Neurophysiol. 1993;86(1):75-7.
125. Jabbari B, Russo MB, Russo ML. Electroencephalogram of asymptomatic adult subjects. Clin Neurophysiol. 2000;111(1):102-5.
126. Eeg-Olofsson O, Petersen I, Sellden U. The development of the electroencephalogram in normal children from the age of 1 through 15 years. Paroxysmal activity. Neuropadiatrie. 1971;2(4):375-404.
127. Borusiak P, Zilbauer M, Jenke AC. Prevalence of epileptiform discharges in healthy children--new data from a prospective study using digital EEG. Epilepsia. 2010;51(7):1185-8. doi: 10.11/j.528-67.2009.02411.x. Epub 2009 Dec 1.
128. Bihege CJ, Langer T, Jenke AC, Bast T, Borusiak P. Prevalence of Epileptiform Discharges in Healthy Infants. J Child Neurol. 2015;30(11):1409-13. doi: 10.177/0883073814565457. Epub 2014 Dec 30.
129. Cavazzuti GB, Cappella L, Nalin A. Longitudinal study of epileptiform EEG patterns in normal children. Epilepsia. 1980;21(1):43-55.
130. Okubo Y, Matsuura M, Asai T, Asai K, Kato M, Kojima T, et al. Epileptiform EEG discharges in healthy children: prevalence, emotional and behavioral correlates, and genetic influences. Epilepsia. 1994;35(4):832-41.
131. Koshino Y, Isaki K. Familial occurrence of the mu rhythm. Clin Electroencephalogr. 1986;17(1):44-50.
132. Bali B, Kull LL, Strug LJ, Clarke T, Murphy PL, Akman CI, et al. Autosomal dominant inheritance of centrotemporal sharp waves in rolandic epilepsy families. Epilepsia. 2007;48(12):2266-72. Epub 007 Jul 28.
133. Degen R, Degen HE, Ahlemeyer K. Contribution to the genetics of symptomatic generalized tonic-clonic seizures: waking and sleep EEGs in siblings. Acta Neurol Scand. 1996;93(1):9-13.
134. van Beijsterveldt CE, Molenaar PC, de Geus EJ, Boomsma DI. Heritability of human brain functioning as assessed by electroencephalography. Am J Hum Genet. 1996;58(3):562-73.

135. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011;2011:156869.(doi):10.1155/2011/156869. Epub 2010 Dec 23.
136. Egli M, Hess R, Kuritzkes G. [The significance of rhythmic mid-temporal discharges (author's transl)]. EEG EMG Z Elektroenzephalogr Elektromyogr Verwandte Geb. 1978;9(2):74-85.
137. Lipman IJ, Hughes JR. Rhythmic mid-temporal discharges. An electro-clinical study. Electroencephalogr Clin Neurophysiol. 1969;27(1):43-7.
138. Lin YY, Wu ZA, Hsieh JC, Yu HY, Kwan SY, Yen DJ, et al. Magnetoencephalographic study of rhythmic mid-temporal discharges in non-epileptic and epileptic patients. Seizure. 2003;12(4):220-5.
139. Koutroumanidis M, Hennessy MJ, Sparkes M, Binnie CD. Continuous bitemporal rhythmic subclinical epileptiform activity in an adult without epileptic seizures. Clin Neurophysiol. 2000;111(4):600-3.
140. Hennessy MJ, Koutroumanidis M, Hughes E, Binnie CD. Psychomotor EEG variant of Gibbs: an association with underlying structural pathology. Clin Neurophysiol. 2001;112(4):686-7.
141. Tye C, Rijsdijk F, McLoughlin G. Genetic overlap between ADHD symptoms and EEG theta power. Brain Cogn. 2014;87:168-72.(doi):10.1016/j.bandc.2014.03.010. Epub Apr 19.
142. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 2014;59(1):5-15. doi: 0.1038/jhg.2013.114. Epub Nov 7.
143. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011;12(11):745-55. doi: 10.1038/nrg3031.
144. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. doi: 10.1093/nar/gkq603. Epub 2010 Jul 3.
145. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012;28(21):2747-54. doi: 10.1093/bioinformatics/bts526. Epub 2012 Aug 31.
146. Makrythanasis P, Antonarakis SE. Pathogenic variants in non-protein-coding sequences. Clin Genet. 2013;84(5):422-8. doi: 10.1111/cge.12272. Epub 2013 Sep 23.
147. Coffey AJ, Kokocinski F, Calafato MS, Scott CE, Palta P, Drury E, et al. The GENCODE exome: sequencing the complete human exome. Eur J Hum Genet. 2011;19(7):827-31. doi: 10.1038/ejhg.2011.28. Epub Mar 2.
148. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet. 2012;91(4):597-607. doi: 10.1016/j.ajhg.2012.08.005.
149. Salgado D, Bellgard MI, Desvignes JP, Beroud C. How to Identify Pathogenic Mutations among All Those Variations: Variant Annotation and Filtration in the Genome Sequencing Era. Hum Mutat. 2016;7(10):23110.
150. de Kovel CG, Brilstra EH, van Kempen MJ, Van't Slot R, Nijman IJ, Afawi Z, et al. Targeted sequencing of 351 candidate genes for epileptic encephalopathy in a large cohort of patients. Mol Genet Genomic Med. 2016;4(5):568-80. doi: 10.1002/mgg3.235. eCollection 2016 Sep.
151. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81. doi: 10.38/nprot.2009.86. Epub Jun 25.
152. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9. doi: 10.1038/nmeth0410-248.

153. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5. doi: 10.1038/ng.2892. Epub 014 Feb 2.
154. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013;9(8):e1003709. doi: 10.1371/journal.pgen.. Epub 2013 Aug 22.
155. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am J Hum Genet. 2014;95(4):360-70. doi: 10.1016/j.ajhg.2014.08.013. Epub Sep 25.
156. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014;42(22):13534-44.
157. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6. doi: 10.1038/nbt.754.
158. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. doi: 10.1038/nature15393.
159. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013;493(7431):216-20. doi: 10.1038/nature11690. Epub 2012 Nov 28.
160. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91. doi: 10.1038/nature19057.
161. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. Hum Mol Genet. 2012;21(R1):R1-9. Epub 2012 Sep 13.
162. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519(7542):223-8. doi: 10.1038/nature14135. Epub 2014 Dec 24.
163. Ran X, Li J, Shao Q, Chen H, Lin Z, Sun ZS, et al. EpilepsyGene: a genetic resource for genes and mutations related to epilepsy. Nucleic Acids Res. 2015;43(Database issue):D893-9. doi: 10.1093/nar/gku943. Epub 2014 Oct 16.
164. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. Genome Res. 2011;21(6):961-73. doi: 10.1101/gr.112326.110. Epub 2010 Oct 27.
165. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;25(10).
166. Lo-Castro A, Curatolo P. Epilepsy associated with autism and attention deficit hyperactivity disorder: is there a genetic link? Brain Dev. 2014;36(3):185-93. doi: 10.1016/j.braindev.2013.04.013. Epub May 29.
167. Lal D, Ruppert AK, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenite D, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet. 2015;11(5):e1005226. doi: 10.1371/journal.pgen.. eCollection 2015 May.
168. Johnson MR, Shkura K, Langley SR, Delahaye-Duriez A, Srivastava P, Hill WD, et al. Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease. Nat Neurosci. 2016;19(2):223-32. doi: 10.1038/nn.4205. Epub 2015 Dec 21.
169. Tuchman R, Moshe SL, Rapin I. Convulsing toward the pathophysiology of autism. Brain Dev. 2009;31(2):95-103. doi: 10.1016/j.braindev.2008.09.009. Epub Nov 8.
170. Li J, Cai T, Jiang Y, Chen H, He X, Chen C, et al. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Mol Psychiatry. 2016;21(2):290-7. doi: 10.1038/mp.2015.40. Epub Apr 7.
171. Hamdan FF, Srour M, Capo-Chichi JM, Daoud H, Nassif C, Patry L, et al. De novo mutations in moderate or severe intellectual disability. PLoS Genet. 2014;10(10):e1004772. doi: 10.1371/journal.pgen.. eCollection 2014 Oct.

172. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380(9854):1674-82. doi: 10.016/S0140-6736(12)61480-9. Epub 2012 Sep 27.
173. Tang S, Pal DK. Dissecting the genetic basis of myoclonic-astatic epilepsy. Epilepsia. 2012;53(8):1303-13.
174. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485(7397):237-41. doi: 10.1038/nature10945.
175. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med. 2012;367(20):1921-9. doi: 10.056/NEJMoa1206524. Epub 2012 Oct 3.
176. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485(7397):246-50. doi: 10.1038/nature10989.
177. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012;485(7397):242-5. doi: 10.1038/nature11011.
178. Klassen T, Davis C, Goldman A, Burgess D, Chen T, Wheeler D, et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell. 2011;145(7):1036-48.
179. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, et al. The real-time polymerase chain reaction. Mol Aspects Med. 2006;27(2-3):95-125. Epub 2006 Feb 3.
180. Medrihan L, Rohlmann A, Fairless R, Andrae J, Doring M, Missler M, et al. Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses. J Physiol. 2009;587(Pt 21):5095-106. doi: 10.1113/jphysiol.2009.178236. Epub 2009 Sep 1.
181. Van Houdt JK, Nowakowska BA, Sousa SB, van Schaik BD, Seuntjens E, Avonce N, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. Nat Genet. 2012;2012 Feb 26;44(4):445-9.
182. Sousa SB, Hennekam RC. Phenotype and genotype in Nicolaides-Baraitser syndrome. Am J Med Genet C Semin Med Genet. 2014;166C(3):302-14. doi: 10.1002/ajmg.c.31409. Epub 2014 Aug 28.
183. Mills PB, Footitt EJ, Mills KA, Tuschl K, Aylett S, Varadkar S, et al. Genotypic and phenotypic spectrum of pyridoxine-dependent epilepsy (ALDH7A1 deficiency). Brain. 2010;133(Pt 7):2148-59. doi: 10.1093/brain/awq143. Epub 2010 Jun 16.
184. van Karnebeek CD, Tiebout SA, Niermeijer J, Poll-The BT, Ghani A, Coughlin CR, 2nd, et al. Pyridoxine-Dependent Epilepsy: An Expanding Clinical Spectrum. Pediatr Neurol. 2016;59:6-12.(doi):10.1016/j.pediatrneurol.2015.12.013. Epub 6 Jan 11.
185. D'Arca D, Zhao X, Xu W, Ramirez-Martinez NC, Iavarone A, Lasorella A. Huwe1 ubiquitin ligase is essential to synchronize neuronal and glial differentiation in the developing cerebellum. Proc Natl Acad Sci U S A. 2010;107(13):5875-80. doi: 10.1073/pnas.0912874107. Epub 2010 Mar 15.
186. Froyen G, Corbett M, Vandewalle J, Jarvela I, Lawrence O, Meldrum C, et al. Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. Am J Hum Genet. 2008;82(2):432-43. doi: 10.1016/j.ajhg.2007.11.002. Epub 8 Jan 24.
187. Orivoli S, Pavlidis E, Cantalupo G, Pezzella M, Zara F, Garavelli L, et al. Xp11.22 Microduplications Including HUWE1: Case Report and Literature Review. Neuropediatrics. 2016;47(1):51-6. doi: 10.1055/s-0035-1566233. Epub 2015 Nov 20.
188. Isrie M, Kalscheuer VM, Holvoet M, Fieremans N, Van Esch H, Devriendt K. HUWE1 mutation explains phenotypic severity in a case of familial idiopathic intellectual disability. Eur J Med Genet. 2013;56(7):379-82. doi: 10.1016/j.ejmg.2013.05.005. Epub May 27.

189. Kurian MA, Meyer E, Vassallo G, Morgan NV, Prakash N, Pasha S, et al. Phospholipase C beta 1 deficiency is associated with early-onset epileptic encephalopathy. Brain. 2010;133(10):2964-70. doi: 10.1093/brain/awq238. Epub 2010 Sep 9.
190. Poduri A, Chopra SS, Neilan EG, Elhosary PC, Kurian MA, Meyer E, et al. Homozygous PLCB1 deletion associated with malignant migrating partial seizures in infancy. Epilepsia. 2012;53(8):e146-50. doi: 10.1111/j.528-67.2012.03538.x. Epub 2012 Jun 12.
191. Ngoh A, McTague A, Wentzensen IM, Meyer E, Applegate C, Kossoff EH, et al. Severe infantile epileptic encephalopathy due to mutations in PLCB1: expansion of the genotypic and phenotypic disease spectrum. Dev Med Child Neurol. 2014;56(11):1124-8. doi: 10.11/dmcn.12450. Epub 2014 Mar 29.
192. Schoonjans AS, Meuwissen M, Reyniers E, Kooy F, Ceulemans B. PLCB1 epileptic encephalopathies; Review and expansion of the phenotypic spectrum. Eur J Paediatr Neurol. 2016;20(3):474-9. doi: 10.1016/j.ejpn.2016.01.002. Epub Jan 13.
193. Cohen BH, Chinnery PF, Copeland WC. POLG-Related Disorders. 1993.
194. Anagnostou ME, Ng YS, Taylor RW, McFarland R. Epilepsy due to mutations in the mitochondrial polymerase gamma (POLG) gene: A clinical and molecular genetic review. Epilepsia. 2016;57(10):1531-45. doi: 10.111/epi.13508. Epub 2016 Aug 24.
195. Kins S, Betz H, Kirsch J. Collybistin, a newly identified brain-specific GEF, induces submembrane clustering of gephyrin. Nat Neurosci. 2000;3(1):22-9.
196. Harvey K, Duguid IC, Alldred MJ, Beatty SE, Ward H, Keep NH, et al. The GDP-GTP exchange factor collybistin: an essential determinant of neuronal gephyrin clustering. J Neurosci. 2004;24(25):5816-26.
197. Lesca G, Till M, Labalme A, Vallee D, Hugonenq C, Philip N, et al. De novo Xq11.11 microdeletion including ARHGEF9 in a boy with mental retardation, epilepsy, macrosomia, and dysmorphic features. Am J Med Genet A. 2011;155A(7):1706-11. doi: 10.002/ajmg.a.34004. Epub 2011 May 27.
198. Marco EJ, Abidi FE, Bristow J, Dean WB, Cotter P, Jeremy RJ, et al. ARHGEF9 disruption in a female patient is associated with X linked mental retardation and sensory hyperarousal. J Med Genet. 2008;45(2):100-5. Epub 2007 Sep 24.
199. Kalscheuer VM, Musante L, Fang C, Hoffmann K, Fuchs C, Carta E, et al. A balanced chromosomal translocation disrupting ARHGEF9 is associated with epilepsy, anxiety, aggression, and mental retardation. Hum Mutat. 2009;30(1):61-8. doi: 10.1002/humu.20814.
200. Bhat G, LaGrave D, Millson A, Herriges J, Lamb AN, Matalon R. Xq11.1-11.2 deletion involving ARHGEF9 in a girl with autism spectrum disorder. Eur J Med Genet. 2016;59(9):470-3. doi: 10.1016/j.ejmg.2016.05.014. Epub May 27.
201. Chioza B, Wilkie H, Nashef L, Blower J, McCormick D, Sham P, et al. Association between the alpha(1a) calcium channel gene CACNA1A and idiopathic generalized epilepsy. Neurology. 2001;56(9):1245-6.
202. Sander T, Toliat MR, Heils A, Becker C, Nurnberg P. Failure to replicate an allelic association between an exon 8 polymorphism of the human alpha(1A) calcium channel gene and common syndromes of idiopathic generalized epilepsy. Epilepsy Res. 2002;49(2):173-7.
203. Zamponi GW, Lory P, Perez-Reyes E. Role of voltage-gated calcium channels in epilepsy. Pflugers Arch. 2010;460(2):395-403. doi: 10.1007/s00424-009-0772-x. Epub 2009 Dec 20.
204. Damaj L, Lupien-Meilleur A, Lortie A, Riou E, Ospina LH, Gagnon L, et al. CACNA1A haploinsufficiency causes cognitive impairment, autism and epileptic encephalopathy with mild cerebellar symptoms. Eur J Hum Genet. 2015;23(11):1505-12. doi: 10.038/ejhg.2015.21. Epub Mar 4.
205. De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies. Am J Hum Genet. 2016;99(2):287-98. doi: 10.1016/j.ajhg.2016.06.003. Epub Jul 28.
206. Sapio MR, Salzmann A, Vessaz M, Crespel A, Lyons PJ, Malafosse A, et al. Naturally occurring carboxypeptidase A6 mutations: effect on enzyme function and association with

epilepsy. J Biol Chem. 2012;287(51):42900-9. doi: 10.1074/jbc.M112.414094. Epub 2012 Oct 26.
207. Salzmann A, Guipponi M, Lyons PJ, Fricker LD, Sapio M, Lambercy C, et al. Carboxypeptidase A6 gene (CPA6) mutations in a recessive familial form of febrile seizures and temporal lobe epilepsy and in sporadic temporal lobe epilepsy. Hum Mutat. 2012;33(1):124-35. doi: 10.1002/humu.21613. Epub 2011 Oct 31.
208. Sapio MR, Vessaz M, Thomas P, Genton P, Fricker LD, Salzmann A. Novel carboxypeptidase A6 (CPA6) mutations identified in patients with juvenile myoclonic and generalized epilepsy. PLoS One. 2015;10(4):e0123180. doi: 10.1371/journal.pone.. eCollection 2015.
209. Allen NM, Conroy J, Shahwan A, Lynch B, Correa RG, Pena SD, et al. Unexplained early onset epileptic encephalopathy: Exome screening and phenotype expansion. Epilepsia. 2016;57(1):e12-7. doi: 0.1111/epi.13250. Epub 2015 Dec 9.
210. Chirala SS, Chang H, Matzuk M, Abu-Elheiga L, Mao J, Mahon K, et al. Fatty acid synthesis is essential in embryonic development: fatty acid synthase null mutants and most of the heterozygotes die in utero. Proc Natl Acad Sci U S A. 2003;100(11):6358-63. Epub 2003 May 8.
211. Carvill GL, Regan BM, Yendle SC, O'Roak BJ, Lozovaya N, Bruneau N, et al. GRIN2A mutations cause epilepsy-aphasia spectrum disorders. Nat Genet. 2013;45(9):1073-6. doi: 10.38/ng.2727. Epub 013 Aug 11.
212. Lesca G, Rudolf G, Bruneau N, Lozovaya N, Labalme A, Boutry-Kryza N, et al. GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. Nat Genet. 2013;45(9):1061-6. doi: 10.38/ng.2726. Epub 013 Aug 11.
213. Lemke JR, Lal D, Reinthaler EM, Steiner I, Nothnagel M, Alber M, et al. Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. Nat Genet. 2013;45(9):1067-72. doi: 10.38/ng.2728. Epub 013 Aug 11.
214. Marwick K, Skehel P, Hardingham G, Wyllie D. Effect of a GRIN2A de novo mutation associated with epilepsy and intellectual disability on NMDA receptor currents and Mg(2+) block in cultured primary cortical neurons. Lancet. 2015;385(Suppl 1):S65. doi: 10.1016/S0140-6736(15)60380-4.
215. Venkateswaran S, Myers KA, Smith AC, Beaulieu CL, Schwartzentruber JA, Majewski J, et al. Whole-exome sequencing in an individual with severe global developmental delay and intractable epilepsy identifies a novel, de novo GRIN2A mutation. Epilepsia. 2014;55(7):e75-9. doi: 10.1111/epi.12663. Epub 2014 Jun 5.
216. Pierson TM, Yuan H, Marsh ED, Fuentes-Fajardo K, Adams DR, Markello T, et al. GRIN2A mutation and early-onset epileptic encephalopathy: personalized therapy with memantine. Ann Clin Transl Neurol. 2014;1(3):190-8.
217. Lim JS, Kim WI, Kang HC, Kim SH, Park AH, Park EK, et al. Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. Nat Med. 2015;21(4):395-400. doi: 10.1038/nm.3824. Epub 2015 Mar 23.
218. Moller RS, Weckhuysen S, Chipaux M, Marsan E, Taly V, Bebin EM, et al. Germline and somatic mutations in the MTOR gene in focal cortical dysplasia and epilepsy. Neurol Genet. 2016;2(6):e118. eCollection 2016 Dec.
219. Otsu M, Hiles I, Gout I, Fry MJ, Ruiz-Larrea F, Panayotou G, et al. Characterization of two 85 kd proteins that associate with receptor tyrosine kinases, middle-T/pp60c-src complexes, and PI3-kinase. Cell. 1991;65(1):91-104.
220. Mirzaa GM, Conti V, Timms AE, Smyser CD, Ahmed S, Carter M, et al. Characterisation of mutations of the phosphoinositide-3-kinase regulatory subunit, PIK3R2, in perisylvian polymicrogyria: a next-generation sequencing study. Lancet Neurol. 2015;14(12):1182-95. doi: 10.016/S474-4422(15)00278-1. Epub 2015 Oct 29.

221. Riviere JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, et al. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. Nat Genet. 2012;44(8):934-40. doi: 10.1038/ng.2331.
222. Terrone G, Voisin N, Abdullah Alfaiz A, Cappuccio G, Vitiello G, Guex N, et al. De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia. Eur J Hum Genet. 2016;24(9):1359-62. doi: 10.038/ejhg.2016.7. Epub Feb 10.
223. Saitsu H, Tohyama J, Kumada T, Egawa K, Hamada K, Okada I, et al. Dominant-negative mutations in alpha-II spectrin cause West syndrome with severe cerebral hypomyelination, spastic quadriplegia, and developmental delay. Am J Hum Genet. 2010;86(6):881-91. doi: 10.1016/j.ajhg.2010.04.013. Epub May 20.
224. Hamdan FF, Saitsu H, Nishiyama K, Gauthier J, Dobrzeniecka S, Spiegelman D, et al. Identification of a novel in-frame de novo mutation in SPTAN1 in intellectual disability and pontocerebellar atrophy. Eur J Hum Genet. 2012;20(7):796-800. doi: 10.1038/ejhg.2011.271. Epub 2 Jan 18.
225. Tohyama J, Nakashima M, Nabatame S, Gaik-Siew C, Miyata R, Rener-Primec Z, et al. SPTAN1 encephalopathy: distinct phenotypes and genotypes. J Hum Genet. 2015;60(4):167-73. doi: 10.1038/jhg.2015.5. Epub Jan 29.
226. Gerber SH, Rah JC, Min SW, Liu X, de Wit H, Dulubova I, et al. Conformational switch of syntaxin-1 controls synaptic vesicle fusion. Science. 2008;321(5895):1507-10. doi: 10.126/science.1163174. Epub 2008 Aug 14.
227. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013;45(7):825-30. doi: 10.1038/ng.2646. Epub 013 May 26.
228. Torkamani A, Bersell K, Jorge BS, Bjork RL, Jr., Friedman JR, Bloss CS, et al. De novo KCNB1 mutations in epileptic encephalopathy. Ann Neurol. 2014;76(4):529-40. doi: 10.1002/ana.24263. Epub 2014 Sep 19.
229. Saitsu H, Akita T, Tohyama J, Goldberg-Stern H, Kobayashi Y, Cohen R, et al. De novo KCNB1 mutations in infantile epilepsy inhibit repetitive neuronal firing. Sci Rep. 2015;5:15199.(doi):10.1038/srep15199.
230. Veeramah KR, Johnstone L, Karafet TM, Wolf D, Sprissler R, Salogiannis J, et al. Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. Epilepsia. 2013;54(7):1270-81.
231. Yang Y, Vasylyev DV, Dib-Hajj F, Veeramah KR, Hammer MF, Dib-Hajj SD, et al. Multistate structural modeling and voltage-clamp analysis of epilepsy/autism mutation Kv10.2-R327H demonstrate the role of this residue in stabilizing the channel closed state. J Neurosci. 2013;33(42):16586-93. doi: 10.1523/JNEUROSCI.2307-13.2013.
232. Steffenburg U, Hagberg G, Hagberg B. Epilepsy in a representative series of Rett syndrome. Acta Paediatr. 2001;90(1):34-9.
233. Christodoulou J, Ho G. MECP2-Related Disorders. 1993.
234. Bianciardi L, Fichera M, Failla P, Di Marco C, Grozeva D, Mencarelli MA, et al. MECP2 missense mutations outside the canonical MBD and TRD domains in males with intellectual disability. J Hum Genet. 2016;61(2):95-101. doi: 10.1038/jhg.2015.118. Epub Oct 22.
235. Acampa M, Guideri F. Cardiac disease and Rett syndrome. Arch Dis Child. 2006;91(5):440-3.
236. Zhu T, Liang C, Li D, Tian M, Liu S, Gao G, et al. Histone methyltransferase Ash1L mediates activity-dependent repression of neurexin-1alpha. Sci Rep. 2016;6:26597.(doi):10.1038/srep26597.
237. Woodage T, Basrai MA, Baxevanis AD, Hieter P, Collins FS. Characterization of the CHD family of proteins. Proc Natl Acad Sci U S A. 1997;94(21):11472-7.
238. Weiss K, Terhal PA, Cohen L, Bruccoleri M, Irving M, Martinez AF, et al. De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. Am J Hum Genet. 2016;99(4):934-41. doi: 10.1016/j.ajhg.2016.08.001. Epub Sep 8.

239. Kim BJ, Kim AR, Lee C, Kim SY, Kim NK, Chang MY, et al. Discovery of CDH23 as a Significant Contributor to Progressive Postlingual Sensorineural Hearing Loss in Koreans. PLoS One. 2016;11(10):e0165680. doi: 10.1371/journal.pone.. eCollection 2016.
240. Palmer EE, Stuhlmann T, Weinert S, Haan E, Van Esch H, Holvoet M, et al. De novo and inherited mutations in the X-linked gene CLCN4 are associated with syndromic intellectual disability and behavior and seizure disorders in males and females. Mol Psychiatry. 2016;23(10):135.
241. Sandford E, Bird TD, Li JZ, Burmeister M. PRICKLE2 Mutations Might Not Be Involved in Epilepsy. Am J Hum Genet. 2016;98(3):588-9. doi: 10.1016/j.ajhg.2016.01.009.
242. Perissinotti PP, Ethington EA, Almazan E, Martinez-Hernandez E, Kalil J, Koob MD, et al. Calcium current homeostasis and synaptic deficits in hippocampal neurons from Kelch-like 1 knockout mice. Front Cell Neurosci. 2014;8:444.(doi):10.3389/fncel.2014.00444. eCollection 2014.
243. Chen WL, Lin JW, Huang HJ, Wang SM, Su MT, Lee-Chen GJ, et al. SCA8 mRNA expression suggests an antisense regulation of KLHL1 and correlates to SCA8 pathology. Brain Res. 2008;1233:176-84.(doi):10.1016/j.brainres.2008.07.096. Epub Aug 3.
244. Tan WH, Bacino CA, Skinner SA, Anselm I, Barbieri-Welge R, Bauer-Carlin A, et al. Angelman syndrome: Mutations influence features in early childhood. Am J Med Genet A. 2011;155A(1):81-90. doi: 10.1002/ajmg.a.33775.
245. Sadikovic B, Fernandes P, Zhang VW, Ward PA, Miloslavskaya I, Rhead W, et al. Mutation Update for UBE3A variants in Angelman syndrome. Hum Mutat. 2014;35(12):1407-17. doi: 10.002/humu.22687.
246. Viani F, Romeo A, Viri M, Mastrangelo M, Lalatta F, Selicorni A, et al. Seizure and EEG patterns in Angelman's syndrome. J Child Neurol. 1995;10(6):467-71.
247. Vendrame M, Loddenkemper T, Zarowski M, Gregas M, Shuhaiber H, Sarco DP, et al. Analysis of EEG patterns and genotypes in patients with Angelman syndrome. Epilepsy Behav. 2012;23(3):261-5. doi: 10.1016/j.yebeh.2011.11.027. Epub 2 Feb 16.
248. Helbig KL, Farwell Hagman KD, Shinde DN, Mroske C, Powis Z, Li S, et al. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet Med. 2016;18(9):898-905. doi: 10.1038/gim.2015.186. Epub 6 Jan 21.
249. De Molfetta GA, Ferreira CA, Vidal DO, Giuliani Lde R, Maldonado MJ, Silva WA, Jr. 1031-1034delTAAC (Leu125Stop): a novel familial UBE3A mutation causing Angelman syndrome in two siblings showing distinct phenotypes. BMC Med Genet. 2012;13:124.(doi):10.1186/471-2350-13-124.
250. Piton A, Redin C, Mandel JL. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. Am J Hum Genet. 2013;93(2):368-83. doi: 10.1016/j.ajhg.2013.06.013. Epub Jul 18.
251. Pittau F, Korff CM, Nordli DR, Jr. Epileptic spasms in epilepsy with myoclonic-atonic seizures (Doose syndrome). Epileptic Disord. 2016;18(3):289-96. doi: 10.1684/epd.2016.0854.
252. Caraballo RH, Fortini S, Reyes G, Carpio Ruiz A, Sanchez Fuentes SV, Ramos B. Epileptic spasms in clusters and associated syndromes other than West syndrome: A study of 48 patients. Epilepsy Res. 2016;123:29-35.(doi):10.1016/j.eplepsyres.2016.03.006. Epub Mar 24.
253. Jeste SS, Tuchman R. Autism Spectrum Disorder and Epilepsy: Two Sides of the Same Coin? J Child Neurol. 2015;30(14):1963-71. doi: 10.177/0883073815601501. Epub 2015 Sep 14.
254. Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science. 2008;320(5880):1224-9. doi: 10.126/science.1153252.
255. Williams AE, Giust JM, Kronenberger WG, Dunn DW. Epilepsy and attention-deficit hyperactivity disorder: links, risks, and challenges. Neuropsychiatr Dis Treat. 2016;12:287-96.(doi):10.2147/NDT.S81549. eCollection 2016.
256. Kwong KL, Lam D, Tsui S, Ngan M, Tsang B, Lam SM. Attention Deficit Hyperactivity Disorder in Adolescents With Epilepsy. Pediatr Neurol. 2016;57:56-63.(doi):10.1016/j.pediatrneurol.2015.12.022. Epub 6 Jan 7.

257. Buelow JM, Perkins SM, Johnson CS, Byars AW, Fastenau PS, Dunn DW, et al. Adaptive functioning in children with epilepsy and learning problems. J Child Neurol. 2012;27(10):1241-9.
258. Matson JL, Bamburg JW, Mayville EA, Khan I. Seizure disorders in people with intellectual disability: an analysis of differences in social functioning, adaptive functioning and maladaptive behaviours. J Intellect Disabil Res. 1999;43(Pt 6):531-9.
259. Sillanpaa M, Helen Cross J. The psychosocial impact of epilepsy in childhood. Epilepsy Behav. 2009;15(Suppl 1):S5-10. doi: .1016/j.yebeh.2009.03.007. Epub May 1.
260. Alfstad KA, Torgersen H, Van Roy B, Hessen E, Hansen BH, Henning O, et al. Psychiatric comorbidity in children and youth with epilepsy: An association with executive dysfunction? Epilepsy Behav. 2016;56:88-94.(doi):10.1016/j.yebeh.2016.01.007. Epub Feb 3.
261. Vega YH, Smith A, Cockerill H, Tang S, Agirre-Arrizubieta Z, Goyal S, et al. Risk factors for reading disability in families with rolandic epilepsy. Epilepsy Behav. 2015;53:174-9.(doi):10.1016/j.yebeh.2015.10.016. Epub Nov 12.
262. Jayalakshmi SS, Mohandas S, Sailaja S, Borgohain R. Clinical and electroencephalographic study of first-degree relatives and probands with juvenile myoclonic epilepsy. Seizure. 2006;15(3):177-83. Epub 2006 Feb 21.
263. Brunnhuber F, Amin D, Nguyen Y, Goyal S, Richardson MP. Development, evaluation and implementation of video-EEG telemetry at home. Seizure. 2014;23(5):338-43. doi: 10.1016/j.seizure.2014.01.009. Epub Jan 20.
264. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, et al. Variation in genome-wide mutation rates within and between human families. Nat Genet. 2011;43(7):712-4. doi: 10.1038/ng.862.
265. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8(8):610-8.
266. Kuroda Y, Ohashi I, Naruto T, Ida K, Enomoto Y, Saito T, et al. Delineation of the KIAA2022 mutation phenotype: two patients with X-linked intellectual disability and distinctive features. Am J Med Genet A. 2015;167(6):1349-53. doi: 10.002/ajmg.a.37002. Epub 2015 Apr 21.
267. Arsov T, Mullen SA, Rogers S, Phillips AM, Lawrence KM, Damiano JA, et al. Glucose transporter 1 deficiency in the idiopathic generalized epilepsies. Ann Neurol. 2012;72(5):807-15. doi: 10.1002/ana.23702.
268. Parrini E, Marini C, Mei D, Galuppi A, Cellini E, Pucatti D, et al. Diagnostic Targeted Resequencing in 349 Patients with Drug-Resistant Pediatric Epilepsies Identifies Causative Mutations in 30 Different Genes. Hum Mutat. 2017;38(2):216-25. doi: 10.1002/humu.23149. Epub 2016 Dec 9.
269. Borjesson SI, Elinder F. Structure, function, and modification of the voltage sensor in voltage-gated ion channels. Cell Biochem Biophys. 2008;52(3):149-74. doi: 10.1007/s12013-008-9032-5. Epub 2008 Nov 7.
270. Ju M, Wray D. Molecular identification and characterisation of the human eag2 potassium channel. FEBS Lett. 2002;524(1-3):204-10.
271. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 1998;393(6683):386-9.
272. Bramswig NC, Ludecke HJ, Alanay Y, Albrecht B, Barthelmie A, Boduroglu K, et al. Exome sequencing unravels unexpected differential diagnoses in individuals with the tentative diagnosis of Coffin-Siris and Nicolaides-Baraitser syndromes. Hum Genet. 2015;134(6):553-68. doi: 10.1007/s00439-015-1535-8. Epub 2015 Feb 28.
273. Ejaz R, Babul-Hirji R, Chitayat D. The evolving features of Nicolaides-Baraitser syndrome - a clinical report of a 20-year follow-up. Clin Case Rep. 2016;4(4):351-5. doi: 10.1002/ccr3.425. eCollection 2016 Apr.
274. Wolff D, Endele S, Azzarello-Burri S, Hoyer J, Zweier M, Schanze I, et al. In-Frame Deletion and Missense Mutations of the C-Terminal Helicase Domain of SMARCA2 in Three

Patients with Nicolaides-Baraitser Syndrome. Mol Syndromol. 2012;2(6):237-44. Epub 2012 Mar 16.
275. Tsurusaki Y, Okamoto N, Ohashi H, Mizuno S, Matsumoto N, Makita Y, et al. Coffin-Siris syndrome is a SWI/SNF complex disorder. Clin Genet. 2014;85(6):548-54. doi: 10.1111/cge.12225. Epub 2013 Jul 23.
276. Son EY, Crabtree GR. The role of BAF (mSWI/SNF) complexes in mammalian neural development. Am J Med Genet C Semin Med Genet. 2014;166C(3):333-49. doi: 10.1002/ajmg.c.31416. Epub 2014 Sep 5.
277. Hargreaves DC, Crabtree GR. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res. 2011;21(3):396-420. doi: 10.1038/cr.2011.32. Epub Mar 1.
278. Yamada T, Yang Y, Hemberg M, Yoshida T, Cho HY, Murphy JP, et al. Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. Neuron. 2014;83(1):122-34. doi: 10.1016/j.neuron.2014.05.039.
279. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209-15. doi: 10.1038/nature13772. Epub 2014 Oct 29.
280. Wang T, Guo H, Xiong B, Stessman HA, Wu H, Coe BP, et al. De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat Commun. 2016;7:13316.(doi):10.1038/ncomms13316.
281. Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF, et al. Quantifying prion disease penetrance using large population control cohorts. Sci Transl Med. 2016;8(322):322ra9. doi: 10.1126/scitranslmed.aad5169.
282. Farwell Hagman KD, Shinde DN, Mroske C, Smith E, Radtke K, Shahmirzadi L, et al. Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. Genet Med. 2016;11(10):95.
283. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. Lancet Neurol. 2017;16(2):135-43. doi: 10.1016/S474-4422(16)30359-3.
284. Howell KB, Harvey AS, Archer JS. Epileptic encephalopathy: Use and misuse of a clinically and conceptually important concept. Epilepsia. 2016;57(3):343-7. doi: 10.1111/epi.13306. Epub 2016 Jan 18.
285. Panayiotopoulos CP. Idiopathic generalized epilepsies: a review and modern approach. Epilepsia. 2005;46(Suppl 9):1-6.
286. Beureau M GP, Dravet C. Epileptic syndromes in infancy, childhood and adolescence (5th edition). 2012.
287. Ottman R. Analysis of genetically complex epilepsies. Epilepsia. 2005;46(Suppl 10):7-14.
288. Suzuki T, Delgado-Escueta AV, Aguan K, Alonso ME, Shi J, Hara Y, et al. Mutations in EFHC1 cause juvenile myoclonic epilepsy. Nat Genet. 2004;36(8):842-9.
289. Pal DK, Evgrafov OV, Tabares P, Zhang F, Durner M, Greenberg DA. BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy. Am J Hum Genet. 2003;73(2):261-70. Epub 2003 Jun 25.
290. Suzuki T, Delgado-Escueta AV, Aguan K, Alonso ME, Shi J, Hara Y, et al. Mutations in EFHC1 cause juvenile myoclonic epilepsy. Nat Genet. 2004;36(8):842-9. Epub 2004 Jul 18.
291. Heinzen EL, Depondt C, Cavalleri GL, Ruzzo EK, Walley NM, Need AC, et al. Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum Genet. 2012;91(2):293-302.
292. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. 2014;13(9):893-903. doi: 10.1016/S474-4422(14)70171-1. Epub 2014 Jul 30.
293. Trump N, McTague A, Brittain H, Papandreou A, Meyer E, Ngoh A, et al. Improving diagnosis and broadening the phenotypes in early-onset seizure and severe developmental

delay disorders through gene panel analysis. J Med Genet. 2016;53(5):310-7. doi: 10.1136/jmedgenet-2015-103263. Epub 2016 Mar 18.
294. Moller RS, Larsen LH, Johannesen KM, Talvik I, Talvik T, Vaher U, et al. Gene Panel Testing in Epileptic Encephalopathies and Familial Epilepsies. Mol Syndromol. 2016;7(4):210-9. Epub 2016 Aug 20.
295. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24. doi: 10.1038/gim.2015.30. Epub Mar 5.
296. Henshall DC, Hamer HM, Pasterkamp RJ, Goldstein DB, Kjems J, Prehn JH, et al. MicroRNAs in epilepsy: pathophysiology and clinical utility. Lancet Neurol. 2016;15(13):1368-76. doi: 10.016/S474-4422(16)30246-0.
297. Roncon P, Soukupova M, Binaschi A, Falcicchia C, Zucchini S, Ferracin M, et al. MicroRNA profiles in hippocampal granule cells and plasma of rats with pilocarpine-induced epilepsy--comparison with human epileptic samples. Sci Rep. 2015;5:14143.(doi):10.1038/srep14143.
298. Panjwani N, Wilson MD, Addis L, Crosbie J, Wirrell E, Auvin S, et al. A microRNA-328 binding site in PAX6 is associated with centrotemporal spikes of rolandic epilepsy. Ann Clin Transl Neurol. 2016;3(7):512-22. doi: 10.1002/acn3.320. eCollection 2016 Jul.
299. Copy number variant analysis from exome data in 349 patients with epileptic encephalopathy. Ann Neurol. 2015;78(2):323-8. doi: 10.1002/ana.24457. Epub 2015 Jul 1.

Appendix A. Information sheets and consent form

King's College Hospital NHS
NHS Foundation Trust

Genetics of Human Epilepsies – Rare Epilepsies

PATIENT INFORMATION SHEET (Adult version 4 – 05/12/13)

You are being invited to take part in a research study. Before you decide to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.
Part 1 tells you the purpose of this study and what will happen if you take part.
Part 2 gives you more detailed information about the conduct of the study.
Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

PART 1

What is the purpose of the study?
The purpose of this study is to find the genes influencing the epilepsies of childhood and adolescence. Differences in genes or DNA (our genetic code that is passed down from our parents) explain the colour of our hair and eyes. Differences in our genes can also explain why some people develop certain diseases, like epilepsy, and others do not. We are studying individuals with epilepsy, as well as their relatives, in order to find the genes that underlie these epilepsies. By finding the genes, we hope to develop new tests to help diagnose epilepsy and also to develop cures.

Why have I been chosen?
You are being asked to participate because you or a member of your family has epilepsy. Even if you are not affected with epilepsy, your contribution helps to pinpoint the genes, by comparing features between people affected with epilepsy and people not affected.

Do I have to take part?
No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. In some instances, a researcher will discuss the study with you and take consent by telephone. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you or your child receive.

What will happen to me if I take part?
If you take part, the investigation and treatment of your child's epilepsy will not be changed in any way. This research study may involve an EEG (brainwave) recording; some tests of hearing and vision done during the EEG; blood samples and saliva samples; some health questionnaires; and some tests of language and reasoning.

Expenses and payments:
There is no charge for this study. You will not receive any payment for taking part in this study.

What do I have to do?

You will be asked to do some or all of the following. Depending on individual circumstances, these procedures may be performed in hospital, at your local phlebotomy service or in your home.

(a) Interview.
You will be asked to participate in completing questionnaires that take about 50-60 minutes. The questions deal with your medical history, including any symptoms of seizures. We will also ask your permission to review your treatment records if you have had seizures or any relevant medical condition. We may also ask you to contact your relatives, if they have relevant symptoms, to enquire if they would be interested to participate in the study. The information from these questionnaires will be available to the researchers in this study.

(b) EEG recording and evoked potentials
An EEG is a painless procedure and is designed to record the activity of brainwaves from the surface of the head. Wires are lightly attached to your scalp using paste and these are connected to the EEG machine. After the recording, the paste is removed. Recordings will last for about 1-3 hours. Sometimes the technician shines a flickering light in front of your face for a minute or so (photic stimulation); at other times s/he may ask you to breathe deeply for a minute (hyperventilation). During the recording, some tones and clicks will be played into the ears through headphones to measure the brainwave response to sounds. All these procedures are designed to bring out certain features of the brainwaves. Often people having an EEG doze off during the recording and this provides useful information too. The EEG is performed in the EEG department of the hospital.

(c) Blood and saliva sample
You will be asked to have blood drawn from which we will extract a sample of DNA (genetic code), RNA (the message copied from DNA), protein and cells. If you choose to have blood drawn, the amount of blood we will draw is approximately 15-20 ml (1.5-2 tablespoons), and the blood draw takes about 5 minutes. A qualified nurse or phlebotomist will draw the blood. There will be a maximum of one or two blood draws.

If you are unable to give a blood sample, you can give a saliva sample instead. You will be asked to spit into a small flask, which takes 5-10 minutes. We will then extract DNA and RNA from the saliva.

The cells of your body contain a molecule called deoxyribonucleic acid, or DNA for short. DNA is passed down from your parents and carries a code, in the form of genes, which determines your physical characteristics, such as the colour of your hair and eyes; and risk for some diseases. Just as differences in our genetic code help explain why we all look different, these differences can also help to explain why some people develop certain diseases and others do not. Differences in our genetic code may also help explain why some drugs are safe and effective for some people but not for others.

The information we get from the DNA is purely experimental and has no diagnostic value. You will not be informed about the results of these DNA experiments. Your DNA will be used to find the genes for epilepsy in the current study, and may be used by the Researcher in later related studies into the cause of epilepsy and developmental disorders. We might want to include your DNA sample in a research project using high-throughput sequencing.

(i) What is high-throughput sequencing?
So far, most scientific studies have only looked at single genes. New technologies now allow for sequencing large parts of the human genome with a single investigation. The human genome comprises all genes of the human body and the entirety of the hereditary information. ‘Sequencing’ refers to the process of reading the sequence of the hereditary information through technical methods.

Studies using high-throughput sequencing are often performed in national or international research collaborations due to the complexity and high costs of these investigations. In this instance, the generated genetic data may be transferred and stored in external research centers, which may be inside or outside the UK. In addition, we might transfer limited clinical information to external research centers such as gender, type of epilepsy, age of onset. Personal and identifying data will never be transferred to other research centers.

(ii) The NIHR Mental Health Biomedical Research Centre Resource Bioresource initiative.
Your DNA, RNA, protein and cells will be included into the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre (BRC) Bioresource. The NIHR is responsible for supporting research and development across the NHS. The Mental Health BRC Bioresource has been designed to create a resource or 'Biobank' for psychiatric/neurological disorders, which will include biological information (e.g. blood, saliva for analyzing genes), combined with other information (e.g. clinical assessments from your child's neurological examination and existing neuroimaging data). The purpose of this Bioresource is to facilitate research that aims to improve our understanding of different neurological disorders and identify new and better treatments.

The Mental Health BRC Bioresource is also linked to and collaborates with a wider national Bioresource initiative involving a number of other Biomedical Research Centers in the country. This is in order to bring together groups of healthy volunteers, patients and relatives who have agreed to be contacted and can be invited to participate in relevant clinical research. Therefore, some BRC Bioresource participants will be invited to participate in relevant clinical research on the basis of their genetic/biological/clinical information. If you are contacted for any follow-up studies, it is up to you to decide whether you would like to participate or not

Your sample will be anonymised, processed and stored at secure laboratories at the Institute of Psychiatry. The BRC Bioresource team will request the sample to be de-anonymised if you are to be invited to participate in follow-up studies.

(d) Neuropsychological testing and speech recording

You will be asked to participate in an evaluation of thinking, language and coordination. This will be a standard evaluation using written, spoken and physical materials, administered by a trained psychologist. During parts of the assessment you may have an EEG recording. The assessment takes about 4 hours. Additional time will be allocated for rest and refreshment. This assessment looks for patterns of brain function that may help in finding genes for epilepsy and is not intended for educational or career guidance. Your child may ask for pause or termination of testing at any time during the procedure.

You may be asked to repeat a series of words and phrases into a recording machine over about 3 minutes. The pattern of your speech will be analyzed later using a computer to look for patterns associated with the function of certain parts of the brain.

What are the other possible disadvantages and risks of taking part?
The procedures involved in the study are routine and involve minor risks only. During the interview some personal questions may be asked that could cause anxiety or stress. You may stop the interview at any time or refuse to answer any of the questions.

The paste used to affix EEG leads to the head may occasionally produce allergic reactions (rash, itch). There is also a small chance, in persons with epilepsy who are sensitive to flickering light, of triggering a

seizure. However, the EEG will be monitored during this procedure and the light can be switched off as soon as first traces of such a response is detected.

The risks of having blood drawn include soreness and bruising at the puncture site, and sometimes there may be discomfort during the procedure. Occasionally people feel lightheaded or faint. There is a small risk of infection whenever blood is drawn. The amount of blood to be taken is not considered to be a significant amount, and is therefore not expected to have any significant risk to you. There are no risks associated with saliva collection.

The risks associated with the cognitive and speech evaluation include only minor stress or boredom.

Although many precautions are taken to safeguard your privacy, there may be unknown risks to privacy related to the fact that your child's biological samples will be kept for an indefinite period of time.

What are the possible benefits of taking part?
There are no direct benefits to you. However, participation might lead to improved understanding of the causes of epilepsy in children in the future.

What if there is a problem?
Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

Will my taking part in the study be kept confidential?
Yes. All the information about your participation in this study will be kept confidential. The details are included in Part 2.

Contact Details:
For further information, please contact the Researcher leading the project:
Professor Deb Pal
Department of Basic and Clinical Neurosciences
Maurice Wohl Neuroscience Institute
Institute of Psychiatry
LONDON SE5 9NU
Telephone: 020 7848 5183

This completes Part 1 of the Information Sheet.
If the information in Part 1 has interested you and you are considering participation, please continue to read the additional information in Part 2 before making any decision.

PART 2

What will happen if I don't want to carry on with the study?

You can withdraw from this research study at any time, even while you are having the tests done. This will not affect your treatment in any way. Even if you withdraw from the study, we would still like to use any information we might already have collected. However, if you want us to destroy the information we have collected, we will. You will need to tell is whether you wish to withdraw from the rare epilepsy study, Bioresource study or both. Biological samples e.g. saliva, blood, blood products, DNA and RNA will be disposed of in accordance with the Human Tissue Act 2004 and King's College London local regulations. Samples which are surplus to the study or no longer needed will also be disposed of in this manner. Samples and any equipment that samples become attached to, e.g. tubes, will be disinfected and placed in separate designated 'Human material clinical waste bags' and sharp safe bins. These bags will be tagged as such and taken to clinical waste bins in the local disposal area of the Institute of Psychiatry and King's College London.

What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak with the researchers who will do their best to answer your questions (020 7848 5435). If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital. In the event that something does go wrong and you are harmed during the research study there are no special compensation arrangements. If you are harmed and this is due to someone's negligence then you may have grounds for a legal action for compensation against King's College Hospital NHS Trust, but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you.

Will my taking part in this study be kept confidential?

Yes. We will collect information about you, which could identify you personally (for example, because the information includes your name or date of birth). We will also collect information about you because we believe it might be relevant to understanding the research. This will include: information about your epilepsy and its treatment, previous scans and EEGs, dates and times of seizures near to the research tests. This information will be stored on computers owned by the hospital and on computers owned by King's College London, the university associated with this hospital. These computers will be securely controlled by the research team, under the direct responsibility of Professor Deb Pal, and no-one outside the team will have access to your information. We will use the information we collect to answer the questions relevant to this research project. In the future it is possible we might have new research questions which could be answered by looking at your information in new ways. We would seek approval from the Research Ethics Committee to use your information for new research projects. If the Research Ethics Committee believed we should contact you again to ask your permission to re-use your information, we will do so. The information we collect in this project will be kept for 10 years and then destroyed securely. The hospital has a duty to ensure research conducted here is of a high standard and auditors from the hospital may need to review any information we hold about you. The auditors will maintain the highest standards of confidentiality. Procedures for handling, processing, storage and destruction of your data are compliant with the Data Protection Act 1998.

Involvement of your doctor

The doctor looking after you in the hospital will be aware of your participation in this research study. With your permission, we will also inform your GP.

What will happen to the results of the research study?

The scientific results of this research study will be published in scientific and medical journals and may be discussed at scientific meetings. You will not be personally identified in any way.

Who is organizing and funding the research?
This research is being organized by Professor Deb Pal and is sponsored by King's College London. The project is funded by the European Union and Medical Research Council. The Bioresource study is funded by the National Institute for Health Research and Professor Tom Craig and Dr Gerome Breen are responsible for this. Neither the research team, nor your doctor, receives any payment if you take part.

Who has reviewed the study?
The Great Ormond Street Hospital/Institute of Child Health Research Ethics Committee (reference number: 09/H0713/76) has reviewed this rare epilepsy study and given a favourable ethical opinion for conduct in the NHS. The South Central Oxford Research Ethics Committee (reference number: 09/H0606/84) has reviewed the Bioresource study and given a favorable ethics opinion.

You will be given a copy of the information sheet and a copy of your signed consent form to keep.

Thank you for considering taking part in this research project, and thank you for taking the time to read the information sheets.

PATIENT INFORMATION SHEET (Parent for Minor version 4, 05/12/13)

Genetics of Human Epilepsies – Rare Epilepsies

Your child is being invited to take part in a research study. Before you decide to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.
Part 1 tells you the purpose of this study and what will happen if you take part.
Part 2 gives you more detailed information about the conduct of the study.
Ask us if there is anything that is not clear or if you would like more information.
Take time to decide whether or not you wish to take part.

PART 1

What is the purpose of the study?
The purpose of this study is to find the genes influencing the epilepsies of childhood and adolescence. Differences in genes or DNA (our genetic code that is passed down from our parents) explain the colour of our hair and eyes. Differences in our genes can also explain why some people develop certain diseases, like epilepsy, and others do not. We are studying individuals with epilepsy, as well as their relatives, in order to find the genes that underlie these epilepsies. By finding the genes, we hope to develop new tests to help diagnose epilepsy and also to develop cures.

Why has my child been chosen?
Your child is being asked to participate because s/he or a member of your family has epilepsy. Even if your child is not affected with epilepsy, their contribution helps to pinpoint the genes, by comparing features between people affected with epilepsy and people not affected.

Does my child have to take part?
No. It is up to you and your child to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. In some instances, a researcher will discuss the study with you and take consent by telephone. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to my child if s/he takes part?
If your child takes part and s/he is a patient at this hospital, the investigation and treatment of his/her epilepsy will not be changed in any way. This research study involves an EEG (brainwave) recording; blood samples and saliva samples; some health questionnaires; and some tests of language and reasoning.

Expenses and payments:
There is no charge for this study. You will not receive any payment for taking part in this study.

What does my child have to do?
S/he will be asked to do some or all of the following. Depending on individual circumstances, these procedures may be performed in hospital, at your local phlebotomy service or in your home.

(a) Interview.
You and your child will be asked to participate in completing questionnaires that take about 50-60 minutes. The questions deal with your child's medical history, including any symptoms of seizures. We will also ask your permission to review your child's treatment records if s/he has had seizures or any

relevant medical condition. We may also ask you to contact your relatives, if they have relevant symptoms, to enquire if they would be interested to participate in the study. The information from these questionnaires will be available to the researchers in this study.

(b) EEG recording and evoked potentials

An EEG is a painless procedure and is designed to record the activity of brainwaves from the surface of the head. Wires are lightly attached to your scalp using paste and these are connected to the EEG machine. After the recording, the paste is removed. Recordings will last for at least 1-2 hours. Sometimes the technician shines a flickering light in front of your child's face for a minute or so (photic stimulation); at other times s/he may ask your child to breathe deeply for a minute (hyperventilation). During the recording, some tones and clicks will be played into the ears through headphones to measure the brainwave responses to sounds. All these procedures are designed to bring out certain features of the brainwaves. Often people having an EEG doze off during the recording, and this provides useful information too. The EEG is performed in the EEG department of the hospital.

(c) Blood and saliva sample

Your child will be asked to have blood drawn from which we will extract a sample of DNA (genetic code), RNA (the message copied from DNA), protein and cells. If you and your child choose to have blood drawn, the amount of blood we will draw is approximately 15-20 ml (1.5-2 tablespoons), and the blood draw takes about 5 minutes. A qualified nurse or phlebotomist will draw the blood. There will be a maximum of one or two blood draws.

If your child is unable to give a blood sample, s/he can give a saliva sample instead. S/he will be asked to spit into a small flask, which takes 5-10 minutes. We will then extract DNA and RNA from the saliva.

The cells of your body contain a molecule called deoxyribonucleic acid, or DNA for short. DNA is passed down from your parents and carries a code, in the form of genes, which determines your physical characteristics, such as the colour of your hair and eyes; and risk for some diseases. Just as differences in our genetic code help explain why we all look different, these differences can also help to explain why some people develop certain diseases and others do not. Differences in our genetic code may also help explain why some drugs are safe and effective for some people but not for others.

The information we get from the DNA is purely experimental and has no diagnostic value. You will not be informed about the results of these DNA experiments. Your DNA will be used to find the genes for epilepsy in the current study, and may be used by the Researcher in later related studies into the cause of epilepsy and developmental disorders. We might want to include your DNA sample in a research project using high-throughput sequencing.

(i) What is high-throughput sequencing?

So far, most scientific studies have only looked at single genes. New technologies now allow for sequencing large parts of the human genome with a single investigation. The human genome comprises all genes of the human body and the entirety of the hereditary information. ‘Sequencing’ refers to the process of reading the sequence of the hereditary information through technical methods.

Studies using high-throughput sequencing are often performed in national or international research collaborations due to the complexity and high costs of these investigations. In this instance, the generated genetic data may be transferred and stored in external research centers, which may be inside or outside the UK. In addition, we might transfer limited clinical information to external research centers such as gender, type of epilepsy, age of onset. Personal and identifying data will never be transferred to other research centers.

(ii) The NIHR Mental Health Biomedical Research Centre Resource Bioresource initiative.
Your child's DNA, RNA, protein and cells will be included into the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre (BRC) Bioresource. The NIHR is responsible for supporting research and development across the NHS. The Mental Health BRC Bioresource has been designed to create a resource or 'Biobank' for psychiatric/neurological disorders, which will include biological information (e.g. blood, saliva for analyzing genes), combined with other information (e.g. clinical assessments from your child's neurological examination and existing neuroimaging data). The purpose of this Bioresource is to facilitate research that aims to improve our understanding of different neurological disorders and identify new and better treatments.

The Mental Health BRC Bioresource is also linked to and collaborates with a wider national Bioresource initiative involving a number of other Biomedical Research Centers in the country. This is in order to bring together groups of healthy volunteers, patients and relatives who have agreed to be contacted and can be invited to participate in relevant clinical research. Therefore, some BRC Bioresource participants will be invited to participate in relevant clinical research on the basis of their genetic/biological/clinical information. If you are contacted for any follow-up studies, it is up to you to decide whether you would like your child to participate or not

Your child's sample will be anonymised, processed and stored at secure laboratories at the Institute of Psychiatry. The BRC Bioresource team will request the sample to be de-anonymised if your child is to be invited to participate in follow-up studies.

(d) Neuropsychological testing and speech recording

Children will be asked to participate in an evaluation of thinking, language and coordination. This will be a standard evaluation using written, spoken and physical materials, administered by a trained psychologist. During parts of the assessment your child may have an EEG recording. The assessment takes about 4 hours. Additional time will be allocated for rest and refreshment. This assessment looks for patterns of brain function that may help in finding genes for epilepsy and is not intended for educational guidance. You or your child may ask for pause or termination of testing at any time during the procedure.

Your child may be asked to repeat a series of words and phrases into a recording machine over about 3 minutes. The pattern of your child's speech will be analyzed later using a computer to look for patterns associated with the function of certain parts of the brain.

What are the other possible disadvantages and risks of taking part?
The procedures involved in the study are routine and involve minor risks only. During the interview some personal questions may be asked that could cause anxiety or stress. You or your child may stop the interview at any time or refuse to answer any of the questions.

The paste used to affix EEG leads to the head may occasionally produce allergic reactions (rash, itch). There is also a small chance, in persons with epilepsy who are sensitive to flickering light, of triggering a seizure. However, the EEG will be monitored during this procedure and the light can be switched off as soon as first traces of such as response are detected.

The risks of having blood drawn include soreness and bruising at the puncture site, and sometimes there may be discomfort during the procedure. Occasionally people feel lightheaded or faint. There is a small risk of infection whenever blood is drawn. The amount of blood to be taken is not considered to be a

significant amount, and is therefore not expected to have any significant risk to your child. There are no risks associated with saliva collection.

The risks associated with the cognitive and speech evaluation include only minor stress or boredom.

Although many precautions are taken to safeguard your child's privacy, there may be unknown risks to privacy related to the fact that your child's biological samples will be kept for an indefinite time.

What are the possible benefits of taking part?
There are no direct benefits to you or your child. However, participation might lead to improved understanding of the causes of epilepsy in children in the future.

What if there is a problem?
Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

Will my taking part in the study be kept confidential?
Yes. All the information about your participation in this study will be kept confidential. The details are included in Part 2.

Contact Details:
For further information, please contact the Researcher leading the project:
Professor Deb Pal
Department of Basic and Clinical Neurosciences
Maurice Wohl Neuroscience Institute
Institute of Psychiatry
LONDON SE5 9NU
Telephone: 020 7848 5183

This completes Part 1 of the Information Sheet.
If the information in Part 1 has interested you and you are considering participation, please continue to read the additional information in Part 2 before making any decision.

PART 2

What will happen if I don't want to carry on with the study?

You can withdraw from this research study at any time, even while you are having the tests done. This will not affect your treatment in any way. Even if you withdraw from the study, we would still like to use any information we might already have collected. However, if you want us to destroy the information we have collected, we will. You will need to tell is whether you wish to withdraw from the rare epilepsy study, Bioresource study or both. Biological samples e.g. saliva, blood, blood products, DNA and RNA will be disposed of in accordance with the Human Tissue Act 2004 and King's College London local regulations. Samples which are surplus to the study or no longer needed will also be disposed of in this manner. Samples and any equipment that samples become attached to, e.g. tubes, will be disinfected and placed in separate designated 'Human material clinical waste bags' and sharp safe bins. These bags will be tagged as such and taken to clinical waste bins in the local disposal area of the Institute of Psychiatry and King's College London.

What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak with the researchers who will do their best to answer your questions (020 7848 5435). If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital. In the event that something does go wrong and you are harmed during the research study there are no special compensation arrangements. If you are harmed and this is due to someone's negligence then you may have grounds for a legal action for compensation against King's College Hospital NHS Trust, but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you.

Will my taking part in this study be kept confidential?

Yes. We will collect information about you, which could identify you personally (for example, because the information includes your name or date of birth). We will also collect information about you because we believe it might be relevant to understanding the research. This will include: information about your epilepsy and its treatment, previous scans and EEGs, dates and times of seizures near to the research tests. This information will be stored on computers owned by the hospital and on computers owned by King's College London, the university associated with this hospital. These computers will be securely controlled by the research team, under the direct responsibility of Professor Deb Pal, and no-one outside the team will have access to your information. We will use the information we collect to answer the questions relevant to this research project. In the future it is possible we might have new research questions which could be answered by looking at your information in new ways. We would seek approval from the Research Ethics Committee to use your information for new research projects. If the Research Ethics Committee believed we should contact you again to ask your permission to re-use your information, we will do so. The information we collect in this project will be kept for 10 years and then destroyed securely. The hospital has a duty to ensure research conducted here is of a high standard and auditors from the hospital may need to review any information we hold about you. The auditors will maintain the highest standards of confidentiality. Procedures for handling, processing, storage and destruction of your data are compliant with the Data Protection Act 1998.

Involvement of your doctor

The doctor looking after you in the hospital will be aware of your participation in this research study. With your permission, we will also inform your GP.

What will happen to the results of the research study?

The scientific results of this research study will be published in scientific and medical journals and may be discussed at scientific meetings. You will not be personally identified in any way.

Who is organizing and funding the research?
This research is being organized by Professor Deb Pal and is sponsored by King's College London. The project is funded by the European Union and Medical Research Council. The Bioresource study is funded by the National Institute for Health Research and Professor Tom Craig and Dr Gerome Breen are responsible for this. Neither the research team, nor your doctor, receives any payment if you take part.

Who has reviewed the study?
The Great Ormond Street Hospital/Institute of Child Health Research Ethics Committee (reference number: 09/H0713/76) has reviewed this rare epilepsy study and given a favourable ethical opinion for conduct in the NHS. The South Central Oxford Research Ethics Committee (reference number: 09/H0606/84) has reviewed the Bioresource study and given a favorable ethics opinion.

You will be given a copy of the information sheet and a copy of your signed consent form to keep.

Thank you for considering taking part in this research project, and thank you for taking the time to read the information sheets.

CONSENT FORM 5/12/13 version 4

Title of Project: Genetics of Human Epilepsies – Rare Epilepsies

Name of Researcher: Professor Deb Pal **Please initial box**

1.	I confirm that I have read and understand the information sheet dated 05/12/13 (version 4) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. Biological samples will be destroyed as detailed.	
3.	I understand that relevant sections of any of my medical notes from hospital and GP medical records and data collected during the study, may be looked at by responsible individuals from regulatory authorities or from the NHS trust, or the BRC Bioresource team, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records	
4.	I agree to have my/child's samples and data included in the National NIHR Bioresource	
5.	I agree that my biological material e.g. blood, saliva, DNA can be removed and used for the above study and that I have been made aware of surplus material disposal methods according to the Human Tissue Act 2004.	
6.	I agree that any biological material surplus to this study and relevant clinical information can be used in future related research, which has been approved by a recognised Research Ethics Committee	
7.	I agree to be contacted in the future by the BRC Bioresource team requesting my/child's participation in similar studies or other medical research studies	
8.	I agree that my blood/tissue sample may be used for high throughput sequencing projects and that my data may be transferred, managed and stored externally. I understand that I will not receive any information about my/child's individual results.	
9.	I agree to my GP being informed of my participation in the study	
10.	I agree to take part in the above study	

Name of Subject	Signature	Date
Name of Parent/Guardian	Signature	Date
Researcher/Consenter	Signature	Date

When completed, 1 for patient; 1 for researcher site file; 1 (original) to be kept in medical notes

Appendix B. Questionnaires

Genetics in Human Epilepsy Study Myoclonic Astatic Epilepsy - Postal Form C_V3

1. Today's Date:

2. Name of person completing form:

3. Your relationship to Child:

4. Your full address:

5. Your contact Home number:
   Mobile number:
   Email:

6. Name of Child:

7. Is your Child: □ Male □ Female

8. Child's birth date:

9. Child's neurologist and Hospital:

---

10. Were there any complications in pregnancy?

    □ No □ Yes, please specify:

11. Was your child born:

    □ Expected dates

    □ Early, please specify at how many weeks your child was born:

    □ Late, please specify at how many weeks your child was born:

12. How was you child born?

□ Normal vaginal delivery

□ Forceps or Ventouse (vacuum) extraction

□ Planned caesarian section

□ Emergency caesarian section

□ Other, please specify:

13. What was the birth weight of your child?

14. What was the condition of your child at the time of birth?

□ Normal, good

□ Other, please specify:

15. How old was your child when s/he could walk without help?

□ Less than 18 months

□ More than 18 months, please specify:

16. Did s/he ever have trouble with any of the following? (Check those that apply):

□ Breast or bottle feeding? (latching; took long time)

□ Chewing solids

□ Sucking through a straw

□ Blowing bubbles or candles

□ Sticking out the tongue

17. In preschool or kindergarten, did you notice that your child:

17.1. Spoke their first words with meaning late, ie had not spoken their first word by 18 months

□ No □ Yes

17.2. Spoke their first sentence late ie only after two and a half years age

□ No □ Yes

17.3. S/he was missing off the consonants at the beginning and/or ends of words, even after two and a half years of age

□ No □ Yes

17.4. His/her words were unintelligible to strangers even after three years of age

□ No □ Yes

17.5. Stuttered after the age of four years

□ No □ Yes

17.6. Got words or parts of words jumbled

□ No □ Yes

17.7. Could not understand spoken language in a sentence by the age of three years

□ No □ Yes

18. Did s/he ever have speech and language therapy?

□ No □ Yes

19. Do you think your child has a learning problem?

□ No □ Yes, please describe:

---

20. How old was your child when s/he had his first seizures?

____ years ____ months

21. How would you describe your child's first seizure?

22. Does you child have other seizures types? Please describe each one:

(1)

(2)

(3)

(4)

23: Did any of the seizures last for at least 30 minutes or did s/he have several seizures without regaining consciousness over at least 30 minutes?

□ No □ Yes □ Not sure

24. What is the longest time your child has ever been seizure free?

25. What medication for epilepsy is your child currently on?

26. What other medication for epilepsy has your child previously tried?

27. When was your child's last seizure?

28. Does/did your child have seizures with fever? □ No □ Yes

29. Is there anyone else in the family (including cousins, aunties, uncles and grandparent) with epilepsy?

□ No □ Yes, please specify how they are related:

30. Is there anyone else in the family (including cousins, aunties, uncles and grandparent) have seizures with fever?

□ No □ Yes, please specify how they are related:

31. Has you child had a brain scan?

□ No, go to question 32

□ Yes

32. Was your child's brain scan normal?

□ Yes

□ No, please specify:

33. Does your child have any medical problems apart from her/his epilepsy?

□ No

□ Yes, please specify:

33. Would you describe your child as having behavioural problems?

□ No

□ Yes, please specify:

Thank you

You have reached the end of the questionnaire

Ethnicity Form (adapted from 2011 Census, Office for National Statistics)

How would you describe your ethnicity?

Tick **ONE** only

White ethnic group

□English/Welsh/Scottish/ Northern Irish

□Gypsy/Irish Traveller

□Irish

□Other White, please specify:

Mixed/Multiple ethnic groups

□White and Black Caribbean

□White and Black African

□White and Asian

□Other Mixed, please specify:

Asian/Asian British

□Indian

□Pakistani

□Bangladeshi

□Chinese

□Other Asian, please specify:

Black/African/Caribbean/Black British

□African

□Caribbean

□Other Black, please specify:

Other ethnic group

□Arab

□Other ethnic group, please specify:

Appendix C. Conner's CBRS subscales

Subscale	Description of subscale components
A	Oppositional: Are likely to break rules, have problems with authority, are easily annoyed.
B	Cognitive Problems/Inattention: Are likely to be inattentive, have organizational problems, have difficulty completing tasks, have concentration problems.
C	Hyperactivity: Have difficulty sitting still for very long, feel restless and impulsive.
D	Anxious-Shy: Have atypical amount of worries and fears; prone to be emotional and sensitive to criticism, anxious in unfamiliar situations, are shy and withdrawn.
E	Perfectionism: Set high goals for themselves, are very fastidious about the way they do things, are obsessive about their work.
F	Social Problems: Are likely to perceive that they have few friends, have low self-esteem and self-confidence, feel socially detached from their peers.
G	Psychosomatic: report atypical amount of aches and pain.
H	Conner's ADHD Index: Identifies children/adolescents 'at risk' for ADHD.
I	CGI Restless-Impulsive: This subscale indicates restlessness, impulsivity and inattentiveness.
J	CGI Emotional Lability: Individuals with high scores on this subscale are prone to more emotional responses/behaviours than is typical.
K	CGI Total: The CGI Score reflects general problematic behaviour.
L	DSM IV Inattentive: High scores indicate an above average correspondence with the DSM IV diagnostic criteria for Inattentive type ADHD.
M	DSM IV Hyperactive-Impulsive: High scores indicate an above average correspondence with the DSM IV diagnostic criteria for Hyperactive-Impulsive type ADHD.
N	DSM IV Total; High scores indicate an above average correspondence to DSM-IV criteria for combined Inattention and Hyperactive-Impulsive type ADHD.

Appendix D. EEG Proforma

Research ID:	
Research site:	
Age:	
Date of EEG:	
Medication:	
State:	Awake / Drowsy / Asleep
Dominant background:	Frequency: Alpha / Beta / Theta / Delta Reactive EO/EC: Symmetrical:
HV response: Yes / No	If Yes: Slowing / Focal spikes / Generalised discharges
Photic stimulation: Yes / No	If Yes: Type I / II / III / IV Range of PS (Hz)
Focal rhythms: Yes / No	if Yes: R / L Alpha / Beta / Theta/ Delta Frontal / Temporal / Parietal / Occipital None
Generalised discharges: Yes / No	If Yes: Spike and wave / Polyspike and wave / Polyspikes / Slow waves
Focal discharges: Yes /No	If Yes: Sharp / Spikes R / L Frontal / Temporal / Parietal / Occipital
Sleep stage:	N1 / N2 / N3 / REM
Benign EE variants: Yes / No	If Yes describe:
Miscellaneous:	

Appendix E. Annotation scripts

Annotation scripts used in annotation analysis of All_epilepsy and Pure_epilepsy variant (see section 6.11).

```
#!/bin/sh
#$ -S /bin/sh
#$ -cwd
#$ -j y
#$ -pe threaded 2

# annotate with respect to genes
/home/exome/bin/annotate_variation.pl --buildver hg19 --splicing_threshold 10 -geneanno /home/shantang/annotations_ana
lysis/inputs/uniqcombinedBED.input.annovar /home/exome/repository/annotation/annovar/humandb/

# return format back to input format
awk 'BEGIN {FS=OFS="\t"} {if ($1 == "exonic;splicing") print $3,$4,$5,$6,$7,$8,substr($1,1,index($1,";")-1),substr($2,
1,index($2,";")-1),$9,$10,$11,$12,$13 ;
                    else if (($1 == "splicing" || $1 == "ncRNA_splicing") && index($2,"(") > 0 ) print $3,$4,$5,$6,$7
,$8,$1,substr($2,1,index($2,"(")-1),$9,$10,$11,$12,$13 ;
                    else print $3,$4,$5,$6,$7,$8,$1,$2,$9,$10,$11,$12,$13 }' /home/shantang/annotations_analysis/inpu
ts/uniqcombinedBED.input.annovar.variant_function > output_1.annovar

# reannotate with respect to genes
/home/exome/bin/annotate_variation.pl --buildver hg19 --splicing_threshold 10 -geneanno output_1.annovar /home/exome/r
epository/annotation/annovar/humandb/

# return format back to input format
awk 'BEGIN {FS=OFS="\t"} {if ((substr($1,0,6) != "exonic") && ($1 != "splicing")) print $3,$4,$5,$6,$7,$8,$9,$10,$9,".
",$11,$12,$13,$14,$15;
                    else if ($1 == "splicing") print $3,$4,$5,$6,$7,$8,$9,$10,$9,$10":"substr($2,index($2,"(")+1,inde
x($2,")")-index($2,"(")-1),$11,$12,$13,$14,$15}' output_1.annovar.variant_function > output_2.annovar
awk 'BEGIN {FS=OFS="\t"} {print $4,$5,$6,$7,$8,$9,$10,$11,$2,$3,$12,$13,$14,$15,$16}' output_1.annovar.exonic_variant_
function >> output_2.annovar

# cross reference with dbSNP142 "with allelic splitting & left-normalization
/home/exome/bin/annotate_variation.pl --buildver hg19 --filter  -dbtype generic -genericdbfile avsnp142.annovar output
_2.annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb/

# return format back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$2,$13,$14,$15,$16,$17}' output_2.annovar.hg19_generi
c_dropped > output_3.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,".",$11,$12,$13,$14,$15}' output_2.annovar.hg19_generic
_filtered >> output_3.annovar


# cross reference with 1000g
/home/exome/bin/annotate_variation.pl --buildver hg19 -filter -dbtype generic -genericdbfile 1KG_2014_10_final.annovar
 output_3.annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb/

# return format back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$2,$14,$15,$16,$17,$18}' output_3.annovar.hg19_ge
neric_dropped > output_4.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,".",$12,$13,$14,$15,$16}' output_3.annovar.hg19_gen
eric_filtered >> output_4.annovar

# cross reference with EVS
/home/exome/bin/annotate_variation.pl --buildver hg19 --filter  -dbtype generic -genericdbfile hg19_esp6500siv2_all.tx
t output_4.annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb/

# return format back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,$2,$15,$16,$17,$18,$19}' output_4.annovar.hg1
9_generic_dropped > output_5.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,$12,".",$13,$14,$15,$16,$17}' output_4.annovar.hg19
_generic_filtered >> output_5.annovar

#ExAC
/home/exome/bin/annotate_variation.pl --buildver hg19 -filter -dbtype generic -genericdbfile hg19_exac03.txt output_5.
annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb/

#format back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,$15,$2,$16,$17,$18,$19,$20}' output_5.annovar
.hg19_generic_dropped > output_6.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,".",$14,$15,$16,$18}' output_5.annovar.hg19
_generic_filtered >> output_6.annovar
```

```
# annotate with respect to genes
/home/exome/bin/annotate_variation.pl --buildver hg19 --splicing_threshold 10 -geneanno /home/shantang/
annotations_analysis/inputs/test.input.annovar /home/exome/repository/annotation/annovar/humandb/


# return format back to input format
awk 'BEGIN {FS=OFS="\t"} {if ($1 == "exonic;splicing") print $3,$4,$5,$6,$7,$8,substr($1,1,index($1,";"
)-1),substr($2,1,index($2,";")-1),$9,$10,$11,$12,$13 ;
                    else if (($1 == "splicing" || $1 == "ncRNA_splicing") && index($2,"(") > 0 ) print
 $3,$4,$5,$6,$7,$8,$1,substr($2,1,index($2,"(")-1),$9,$10,$11,$12,$13 ;
                    else print $3,$4,$5,$6,$7,$8,$1,$2,$9,$10,$11,$12,$13 }' /home/shantang/annotation
s_analysis/inputs/test.input.annovar.variant_function > output_1.annovar

#CADD
/home/exome/bin/annotate_variation.pl --buildver hg19 --filter  -dbtype generic -genericdbfile hg19_cad
dgt10.txt output_1.annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb/

#back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,$15,$16,$2,$17,$18,$19,$20,$21
}' output_1.annovar.hg19_generic_dropped > output_2.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,".",$15,$16,$17,$18,$19}
' output_1.annovar.hg19_generic_filtered >> output_2.annovar

#SIFT
/home/exome/bin/annotate_variation.pl --buildver hg19 --filter  -dbtype generic -genericdbfile hg19_ljb
26_sift.txt output_2.annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb/

#back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,$15,$16,$2,$17,$18,$19,$20,$21
,$22}' output_2.annovar.hg19_generic_dropped > output_3.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,".",$15,$16,$17,$18,$19,
$20}' output_2.annovar.hg19_generic_filtered >> output_3.annovar

#pp2hvar
/home/exome/bin/annotate_variation.pl --buildver hg19 --filter  -dbtype generic -genericdbfile hg19_ljb
26_pp2hvar.txt output_3.annovar /home/exome/repository/annotation/annovar/humandb/annotation_v2/humandb
/

#back to input format
awk 'BEGIN {FS=OFS="\t"} {print $3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,$15,$16,$2,$17,$18,$19,$20,$21
,$22,$23}' output_3.annovar.hg19_generic_dropped > test_final_output.annovar
awk 'BEGIN {FS=OFS="\t"} {print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,$12,$13,$14,".",$15,$16,$17,$18,$19,
$20,$21}' output_3.annovar.hg19_generic_filtered >> test_final_output.annovar

#remove excess intermediate files
#rm output_*.annovar
```

Appendix F. DNA extraction protocols

Blood DNA extraction protocol

The QIAamp Blood Maxi Kit (Spin Protocol) was used. Volumes stated in the procedure are based on a 5ml blood collection.

Procedure:

- Pipet 500µl of QIAGEN Protease into a 50ml tube. Add the blood collected and mix briefly.
- Add 6ml of Buffer AL and mix by inverting the tube 15 times followed by vigorous shaking for at least 1 minute.
- Incubate at 70^{o}C for 10 minutes.
- Add 5mls of 100% ethanol and mix by inverting the tube 10 times followed by vigorous shaking.
- Transfer the solution taking care not to moisten the rim onto the QIAamp Maxi column placed in a 50ml centrifuge tube. Close the cap and centrifuge at 3000rpm for 3 minutes.
- Remove the QIAamp Maxi column, discard the filtrate and place the QIAamp Maxi column back into the 50ml centrifuge tube.
- Without moistening the rim, add 5ml Buffer AW1 to the QIAamp Maxi column. Close the cap and centrifuge at 5000rpm for 2 minutes.
- Without moistening the rim, add 5ml Buffer AW2 to the QIAamp Maxi column. Close the cap and centrifuge at 5000rpm for 15 minutes.
- Place the QIAamp Maxi column in a clean 50ml centrifuge tube and discard the tube containing the filtrate.
- Pipet 600µl of Buffer AE directly onto the membrane of the QIAamp Maxi column, close the cap and incubate at room temperature for 5 minutes, then centrifuge at 5000rpm for 2 minutes
- Reload the 600µl eluate onto the membrane of the QIAamp Maxi column and then add 500µl Buffer AE.
- Close the cap and incubate at room temperature for 5 minutes then centrifuge at 4500rpm for 5 minutes.
- Check DNA concentrations and A260/A280 wavelength ratio.

Saliva DNA extraction protocol for Oragene OG 250, OG 500 and OG 575

The PrepIT L2P kit was used.

Procedure:

- Transfer entire sample including lid content to 15ml tube.
- Incubate in 50^{o}C water bath for 60 minutes.
- Add 1/25th of sample volume of PrepIT L2P to a 15ml tube and vortex. Eg. For subject 10603, 4.7mls of saliva was obtained and 188ul of PrepIT L2P added.
- Incubate on ice for 10 minutes.
- Centrifuge at 4500rpm for 10 minutes at room temperature.
- Transfer the clear supernatant to a fresh 15ml tube. Discard the waste pellet.
- Add 1.2x of sample volume of 100% ethanol and invert 10 times. For example, for subject 10603, 5.0ml of 100% ethanol was added.
- Leave to stand for 10 minutes.
- Centrifuge at 4500rpm for 10 minutes at room temperature.
- Carefully remove the supernatant being careful not to disturb the pellet. Use a 5ml pipette for most of the supernatant followed by a 200µl or 20µl pipette.
- Add 1ml of freshly made 70% ethanol.
- Leave to stand for 1 minute.
- Completely remove the ethanol and leave to dry inverted for 10 minutes. Ethanol drops on the inner surface of the tube may be removed with tissue paper.
- Rehydrate by adding 800µl of TE buffer. Vortex for 30 seconds being careful to dislodge and mix the pellet.
- Leave to stand for 24 hours then check DNA concentration and A260/280 wavelength ratio.

Saliva DNA extraction protocol for Isohelix GFX-02 4ml GeneFix

The Isohelix DNA extraction protocol uses GSPN-12 GeneFix Full DNA Isolation Kit.

Procedure:

- Vortex the GeneFix saliva collection tube to mix well.
- Add 40µl Proteinase K solution, vortex to mix then incubate at 60^{o} C for 30 minutes.
- Transfer the solution to a 15ml centrifuge tube.
- Add 4ml of SPN buffer, vortex to mix well.
- Centrifuge at 4400rpm for 30 minutes.

- Pour off the supernatant and then re-spin at 3000rpm for 2 minutes.
- Remove all the remaining liquid with a pipette tip without disturbing the DNA pellet.
- Add 400µl TE buffer to the tube, vortex well and leave at room temperature for 5 minutes to re-hydrate the DNA.
- Transfer the sample to a 1.5ml tube and centrifuge at maximum speed 13.4K rpm for 20 minutes at 20^{o}C.
- Check DNA concentration and A260/280 wavelength ratio.

Appendix G. DNA sample IDs

UK DNA sample IDs

Proband ID	Exome sequencing ID	Sex	DNA of family members
3000301	S1040	Male	F, M, B, C
3001301		Male	
3002301	S1271	Male	F, M, S, B, B
3003301	S1272	Female	F, M
3004301	S1273	Male	F, M
00500*	S2286	Male	M
00504*±	S2340	Female	F, M
00505*±	S2341	Male	F, M, B, S
00506*	S2388	Male	F, M
00512*	S2389	Male	F, M
00513*	-	Male	F, M, S
30513	S2574	Female	
00514*	S2287	Male	
00518*	S1616	Male	F, M, B, S
00523*	S2288	Male	F, M
00524*		Male	F, M, B
00525*	S2391	Female	F, M
00526*	S1883	Female	M, B
00528*	S2488	Female	F, M, S
00529*	S2289	Female	F, M
00530*	S2392	Male	F, M
00533*	S2393	Female	F, M
00534*	S2290	Male	F, M
00536		Male	F, M
00538*		Male	M
00539*	S2291	Male	F, M, B
00545*	S2292	Male	F, M, S
00546*	S2342	Female	F, M
00550*	S2343	Male	F, M
00551*	S2293	Female	F, M, S
00554*	S2294	Female	F, M, B
00556*		Male	M
00559*	S2331	Female	F, M
00560*	S2295	Male	F, M
00561*	S2332	Female	F, M
00562*		Male	F, M, C
00566	S2296	Male	F, M
00567		Male	F, M, B

00568*	S2297	Male	F, M
00570*	S2298	Female	F, M
00571	S2390	Male	F, M, B
00572*	S2330	Male	M
00574		Male	F, M, S
00576	S2299	Male	F, M
00577		Male	F, M
00580	S2534	Male	F, M
00584	S2535	Male	M
00586		Male	F, M
00587		Female	F, M
00590		Male	F, M
00591		Female	F, M
00594	S2536	Female	F, M
00595	S2537	Male	F, M
00597		Female	F, M
00598	S2538	Male	F, M
00599	S2539	Male	F, M
00600	S2540	Female	M
00602	S2541	Male	F, M, B, B, S
00603	S2542	Male	F, M
00605		Female	F, M
00607			F, M
00608	S2543	Male	F, M
00609	S2544	Male	M
00610	S2545	Male	F, M
00613		Male	F, M
00615		Male	F, M
00616		Male	F, M
00618		Male	F, M
00621		Male	F, M
00622		Female	F, M
00623		Male	F, M
00625		Male	F, M

*DNA processed and stored at SGDP Biobank, F father, M mother, B brother, S sister.

Italian DNA sample IDs

Proband ID	Exome sequencing ID	Sex
291J	S1394	Female
009Z	S1392	Male
295D	S1395	Female
072G	S1393	Male
138J	S1388	Female
729H	S1390	Male
731J	S1391	Female
660M	S1389	Male
285D	S1384	Female
565D	S1387	Male
347H	S1385	Female
401J	S1386	Male
561D	S1382	Female
262N	S1381	Female
224D	S1380	Female
768H	S1383	Male
204D	S1377	Male
027A	S1376	Male

Appendix H. Gene sets

Neuropsychiatric gene set

A2M	*AIFM3*	*ARID1B*	*BIRC6*	*C9orf30*	*CD247*
AADACL2	*AIP*	*ARIH1*	*BMP1*	*C9orf68*	*CD274*
AAK1	*AK5*	*ARMC3*	*BNC2*	*CA5B*	*CD300LF*
AARS2	*AK7*	*ARMC9*	*BPIFB6*	*CACHD1*	*CD3EAP*
ABCA1	*AKAP6*	*ARNT*	*BRAF*	*CACNA1A*	*CD40*
ABCA12	*AKNA*	*ARNT2*	*BRCA2*	*CACNA1D*	*CD58*
ABCA13	*AKR1C4*	*ARRB2*	*BRD1*	*CACNA1E*	*CD72*
ABCA2	*AKR7A3*	*ARRDC1*	*BRSK2*	*CACNA1G*	*CDADC1*
ABCA4	*AL354898.1*	*ARRDC4*	*BRWD1*	*CACNA1H*	*CDAN1*
ABCB5	*ALDH1A3*	*ASAH2*	*BSPRY*	*CACNA2D2*	*CDC25B*
ABCB9	*ALDH1B1*	*ASAP2*	*BTBD9*	*CACNG3*	*CDC42BPA*
ABCC12	*ALDH1L1*	*ASB1*	*BTN1A1*	*CADPS*	*CDC42BPB*
ABCC3	*ALDH3B2*	*ASB16*	*BTNL8*	*CALML4*	*CDCA7L*
ABCC8	*ALDH9A1*	*ASH1L*	*C10orf11*	*CAMK2A*	*CDH2*
ABCC9	*ALG13*	*ASPH*	*C10orf90*	*CAMK2B*	*CDH23*
ABCD2	*ALK*	*ASRGL1*	*C11orf41*	*CAMK2G*	*CDH5*
ABCG2	*ALMS1*	*ASTE1*	*C12orf28*	*CAMK4*	*CDHR5*
ABI3BP	*ALOX12B*	*ASXL1*	*C12orf41*	*CAMSAP3*	*CDK13*
ABL2	*ALS2CL*	*ATAD2B*	*C12orf51*	*CAMTA1*	*CDK5RAP1*
AC132216.1	*AMPD2*	*ATF7*	*C12orf72*	*CANT1*	*CDKAL1*
ACACA	*AMY2B*	*ATF7IP2*	*C14orf101*	*CAP1*	*CDKL4*
ACACB	*ANK2*	*ATG10*	*C14orf4*	*CAPN10*	*CDKL5*
ACAP2	*ANK3*	*ATIC*	*C15orf38-AP3S2*	*CAPN14*	*CDON*
ACHE	*ANKRD11*	*ATM*	*c15orf40*	*CAPN8*	*CDS2*
ACLY	*ANKRD12*	*ATN1*	*C15orf44*	*CAPN9*	*CDX2*
ACO1	*ANKRD17*	*ATP10B*	*C15orf62*	*CAPRIN2*	*CDYL2*
ACOT4	*ANKRD35*	*ATP11C*	*C16orf62*	*CAPZA1*	*CEACAM21*
ACP2	*ANKRD50*	*ATP1A1*	*C17orf47*	*CARD14*	*CEACAM4*
ACPT	*ANKS1A*	*ATP1A2*	*C17orf53*	*CARKD*	*CEACAM6*
ACTC1	*ANTXR1*	*ATP1A3*	*C18orf25*	*CASC1*	*CECR1*
ACTL6B	*AOC2*	*ATP2A2*	*C18orf26*	*CASK*	*CELA1*
ACTN1	*AOC3*	*ATP2B1*	*C18orf63*	*CASKIN2*	*CELA3B*
ACTN4	*AP2M1*	*ATP2B4*	*C19orf40*	*CASP7*	*CELSR1*
ACTRT2	*AP3B1*	*ATP6V1B2*	*C19orf57*	*CASP9*	*CENPF*
ADAM22	*AP3B2*	*ATP7B*	*C1GALT1*	*CBFA2T3*	*CENPW*
ADAM30	*AP3M1*	*ATP8A1*	*C1orf116*	*CBL*	*CEP192*
ADAM33	*AP4M1*	*ATRX*	*C1orf123*	*CBLL1*	*CEP350*
ADAMTS9	*APAF1*	*ATXN2L*	*C1orf125*	*CBX5*	*CEP55*
ADAMTSL1	*APH1A*	*ATXN3*	*C1orf141*	*CCBE1*	*CEP63*
ADC	*API5*	*AUTS2*	*C1orf173*	*CCDC125*	*CEP85*
ADCK5	*APLF*	*B3GNT4*	*C1QTNF6**	*CCDC132*	*CEP95*
ADCY2	*APLNR*	*B4GALNT3*	*C20orf111*	*CCDC14*	*CGNL1*
ADCY3	*APLP1*	*B4GALNT4*	*C20orf195*	*CCDC15*	*CHAF1A*
ADCY5	*APOC3*	*BACE1*	*c20orf26*	*CCDC164*	*CHAMP1*
ADD1	*APOL4*	*BAI2*	*C20orf26*	*CCDC18*	*CHD1*
ADD3	*ARAP1*	*BARD1*	*C2orf42*	*CCDC36*	*CHD2*
ADGB	*ARFGAP3*	*BAZ1A*	*C3*	*CCDC39*	*CHD3*
ADH4	*ARFGEF1*	*BAZ2B*	*C3orf22*	*CCDC40*	*CHD4*
ADNP	*ARFGEF2*	*BBX*	*C3orf38*	*CCDC78*	*CHD7*
AFAP1	*ARHGAP15*	*BCAM*	*C3orf77*	*CCDC88A*	*CHD8*
AFF3	*ARHGAP19*	*BCAS1*	*C4BPA*	*CCDC89*	*CHEK1*
AFF4	*ARHGAP20*	*BCAT1*	*C4orf37*	*CCDC90B*	*CHIA*
AGAB	*ARHGAP25*	*BCL11A*	*C4orf50*	*CCND1*	*CHMP2A*
AGAP1	*ARHGAP26*	*BCL9L*	*C5orf42*	*CCND3*	*CHRM4*
AGBL5	*ARHGAP31*	*BCLAF1*	*C6orf10*	*CCNJ*	*CHRNA1*
AGER	*ARHGAP35*	*BCOR*	*C6orf130*	*CCNL2*	*CHRNA10*
AGK	*ARHGAP40*	*BCORL1*	*C6orf174*	*CCNT1*	*CHRNA4*
AGPAT3	*ARHGEF10L*	*BEND3*	*C7orf43*	*CCT2*	*CHRND*
AHCY	*ARHGEF26*	*BEST2*	*C7orf59*	*CD101*	*CIAO1*
AHDC1	*ARHGEF9*	*BFSP1*	*C9orf129*	*CD151*	*CISH*
AHNAK2	*ARID1A*	*BHLHE40*	*C9orf144B*	*CD244*	*CIT*

CKAP2L	*CTSB*	*DICER1*	*EED*	*FAM69B*	*GABRA1*
CLCA1	*CTSL1*	*DIP2C*	*EEF1A2*	*FAM76B*	*GABRA3*
CLCN4	*CTSZ*	*DIS3L*	*EFCAB4A*	*FAM83H*	*GABRB1*
CLCN6	*CTTNBP2*	*DIS3L2*	*EFCAB8*	*FAM86C1*	*GABRB2*
CLPX	*CTTNBP2NL*	*DISC1*	*EFHD2*	*FAM8A1*	*GABRB3*
CLRN2	*CUBN*	*DKFZP667F0711*	*EFR3A*	*FAM98A*	*GAD1*
CLRN3	*CUL2*	*DKKL1*	*EFS*	*FANCD2*	*GAK*
CLSTN3	*CUL3*	*DLEC1*	*EGFEM1P*	*FARS2*	*GALC*
CLTC	*CUL5*	*DLG4*	*EGR1*	*FASN*	*GALNT8*
CLTCL1	*CUL9*	*DLG5*	*EHBP1*	*FASTKD2*	*GAPVD1*
CMIP	*CUX2*	*DLGAP1*	*EHD2*	*FAT1*	*GAS2*
CNGA3	*CXorf36*	*DMD*	*EHMT1*	*FBL*	*GAS2L1*
CNKSR2	*CXXC11*	*DMRT3*	*EIF2C1*	*FBLN7*	*GAS2L2*
CNN2	*CYB5A*	*DMXL1*	*EIF3G*	*FBXL19*	*GATAD2B*
CNNM4	*CYFIP2*	*DNAAF2*	*EIF3L*	*FBXL5*	*GBP6*
CNOT1	*CYP1A2*	*DNAH10*	*EIF4G1*	*FBXL6*	*GCKR*
CNOT2	*CYP20A1*	*DNAH11*	*EIF4G2*	*FBXO10*	*GCM1*
CNOT3	*CYP26A1*	*DNAH12*	*EIF5A2*	*FBXO15*	*GCM2*
CNOT4	*CYP26B1*	*DNAH17*	*EIF6*	*FBXO16*	*GET4*
CNOT8	*CYP27C1*	*DNAH5*	*ELK1*	*FBXO18*	*GFOD1*
CNST	*CYP2U1*	*DNAH6*	*ELL*	*FBXO21*	*GFOD2*
CNTN5	*CYP4F3*	*DNAH7*	*EMILIN3*	*FBXO22*	*GGA2*
CNTNAP4	*CYP51A1*	*DNAH8*	*EN2*	*FBXO28*	*GGNBP2*
COBRA1	*CYTH4*	*DNAH9*	*ENAH*	*FBXW9*	*GHR*
COL11A1	*D2HGDH*	*DNAI2*	*ENO2*	*FCGR2B*	*GINS2*
COL11A2	*DAAM1*	*DNAJB9*	*ENOX1*	*FCRL6*	*GIPC2*
COL12A1	*DAB2*	*DNAJC1*	*EP300*	*FETUB*	*GIT1*
COL15A1	*DAG1*	*DNAJC6*	*EPHB1*	*FEZF2*	*GJA10*
COL25A1	*DALRD3*	*DNM1*	*EPHB2*	*FGD3*	*GJA8*
COL2A1	*DAO*	*DNMT1*	*EPPK1*	*FGF12*	*GLB1*
COL3A1	*DBH*	*DNMT3A*	*EPS8L3*	*FGF18*	*GLB1L3*
COL4A2-AS2	*DBNL*	*DNPEP*	*EPX*	*FGF22*	*GLDC*
COL4A3BP	*DBR1*	*DOC2A*	*ERCC6L*	*FGF5*	*GLE1*
COL4A4	*DCAF4*	*DOCK3*	*ERMP1*	*FGFR2*	*GLIS1*
COL5A1	*DCAF7*	*DOCK6*	*ERN2*	*FGFR3*	*GLIS3*
COL5A3	*DCHS2*	*DOCK9*	*ERV3*	*FHDC1*	*GLMN*
COL7A1	*DCX*	*DOM3Z*	*ESPL1*	*FIGNL1*	*GLP1R*
COL8A1	*DDAH1*	*DOT1L*	*ETF1*	*FKBP15*	*GLP2R*
COLQ	*DDI2*	*DPP3*	*ETFB*	*FLG*	*GLRA2*
COQ3	*DDR2*	*DPP4*	*EVC2*	*FLNA*	*GMFG*
COX11	*DDX20*	*DPP7*	*EVI5*	*FLNC*	*GNA11*
CPNE4	*DDX39B*	*DPP9*	*EWSR1*	*FLT4*	*GNAI1*
CPSF4	*DDX3X*	*DPYSL5*	*EXOC2*	*FMN2*	*GNAO1*
CPZ	*DDX50*	*DRD2*	*EXOC7*	*FN1*	*GNAS*
CR2	*DDX58*	*DSCR4*	*EXOSC2*	*FNDC7*	*GNAZ*
CRB1	*DDX60*	*DSG2*	*EXOSC5*	*FNIP1*	*GNL3L*
CREB1	*DDX60L*	*DSP*	*F2*	*FOXG1*	*GNPTAB*
CREBBP	*DEAF1*	*DST*	*F7*	*FOXK2*	*GNRHR*
CRTAC1	*DENND4C*	*DTWD2*	*FADS3*	*FOXN4*	*GOLGA3*
CRY2	*DENND5A*	*DTX4*	*FAF2*	*FOXP1*	*GOLGA5*
CSDE1	*DENR*	*DTYMK*	*FAM102A*	*FOXP2*	*GOLGB1*
CSMD1	*DEPDC7*	*DUS1L*	*FAM116B*	*FPGT-TNNI3K*	*GON4L*
CSMD2	*DET1*	*DUSP14*	*FAM133B*	*FRAS1*	*GPAM*
CSNK1E	*DGCR14*	*DUSP15*	*FAM172A*	*FRAT2*	*GPC1*
CSNK2A1	*DGCR8*	*DUSP3*	*FAM178A*	*FREM2*	*GPR108*
CSPP1	*DGKH*	*DYNC1H1*	*FAM19A2*	*FREM3*	*GPR114*
CST9	*DHDDS*	*DYNC1LI2*	*FAM20B*	*FRMD8*	*GPR124*
CSTF1	*DHDH*	*DYRK1A*	*FAM21C*	*FRMPD1*	*GPR125*
CTBP1	*DHRS4L1*	*DYRK1B*	*FAM45A*	*FRY*	*GPR128*
CTBP2	*DHTKD1*	*EBAG9*	*FAM47E*	*FRYL*	*GPR139*
CTC1	*DHX30*	*EBF3*	*FAM50A*	*FSHR*	*GPR146*
CTCF	*DHX32*	*ECEL1*	*FAM59B*	*FXYD2*	*GPR15*
CTNNA2	*DHX36*	*ECH1*	*FAM5C*	*FYCO1*	*GPR84*
CTNNB1	*DHX40*	*ECHDC2*	*FAM63B*	*G3BP2*	*GPR98*
CTRB1	*DIAPH3*	*ECT2L*	*FAM65C*	*GABBR2*	*GPRASP1*

GPRIN3	*HOXC10*	*KAT6B*	*KLHL32*	*LRRK1*	*MLL2*
GPS1	*HPD*	*KATNAL2*	*KLHL36*	*LTA*	*MLL3*
GPX8	*HRG*	*KBTBD12*	*KLK3*	*LTBP1*	*MLL4*
GRAMD1A	*HRH4*	*KCNA3*	*KLRD1*	*LUC7L3*	*MLLT4*
GRAMD2	*HS2ST1*	*KCNB1*	*KMT2B*	*LY6G6F*	*MMP19*
GRB14	*HSF2*	*KCNC4*	*KRBA1*	*LY75-CD302*	*MMP27*
GREB1L	*HSF2BP*	*KCNE2*	*KRT15*	*LY9*	*MMP8*
GRHL3	*HSPA9*	*KCNG2*	*KRT27*	*LYPD2*	*MMP9*
GRIA1	*HSPB2*	*KCNG3*	*KRT32*	*LZTFL1*	*MMRN2*
GRID1	*HSPB3*	*KCNH1*	*KRT34*	*LZTR1*	*MN1*
GRIK5	*HTATIP2*	*KCNH4*	*KRT77*	*MACC1*	*MOBKL3*
GRIN1	*HUWE1*	*KCNJ4*	*KRT8*	*MADD*	*MPDZ*
GRIN2A	*IFT172*	*KCNK12*	*KRT80*	*MAEA*	*MPHOSPH8*
GRIN2B	*IGBP1*	*KCNK3*	*KRTAP1-3*	*MANSC1*	*MPP6*
GRIN3B	*IGDCC4*	*KCNMA1*	*KRTAP19-1*	*MAP3K1*	*MR1*
GRINL1A	*IGF2R*	*KCNQ2*	*KRTAP19-2*	*MAP3K7*	*MRPL40*
GRIP1	*IGHV3-9*	*KCNQ3*	*KRTAP9-3*	*MAP3K9*	*MRPL44*
GRM1	*IGLV3-27*	*KCNS1*	*L1TD1*	*MAP4*	*MRPS23*
GRM7	*IGSF3*	*KCNT1*	*L3MBTL1*	*MAPK11*	*MRPS27*
GSPT2	*IKBIP*	*KDELR1*	*LAMA1*	*MAPK8IP1*	*MRS2*
GSTT1	*IL12RB2*	*KDELR3*	*LAMA4*	*MAPKBP1*	*MS4A1*
GTF2H4	*IL17RC*	*KDM1A*	*LAMB1*	*MARK1*	*MSANTD1*
GTF2IRD1	*IL18R1*	*KDM1B*	*LAMB2*	*MARK2*	*MSH6*
GTPBP8	*IL1R2*	*KDM3B*	*LAMC1*	*MAST1*	*MSL2*
GZMH	*IL20RB*	*KDM4B*	*LARP6*	*MAST2*	*MSL3*
H2AFV	*ILK*	*KDM6A*	*LARP7*	*MAT1A*	*MSLNL*
H2AFY	*ILVBL*	*KHK*	*LARS*	*MATN1*	*MST1R*
H3F3A	*IMPDH2*	*KIAA0087*	*LASS6*	*MB21D2*	*MTA1*
H6PD	*INCENP*	*KIAA0100*	*LCE1A*	*MBD5*	*MTF1*
HADHA	*INMT*	*KIAA0182*	*LCMT2*	*MCAM*	*MTMR12*
HAPLN4	*INO80*	*KIAA0226*	*LCN10*	*MCF2L*	*MTMR2*
HBS1L	*INPP5B*	*KIAA0317*	*LDHA*	*MCL1*	*MTMR3*
HCK	*INPP5J*	*KIAA0319*	*LDLRAD1*	*MCM2*	*MTMR4*
HCN1	*INSRR*	*KIAA0368*	*LDLRAD3*	*MCPH1*	*MTMR9*
HCN4	*INTS10*	*KIAA0556*	*LEF1*	*MDH1B*	*MTOR*
HDAC4	*INTU*	*KIAA0913*	*LEMD2*	*MDM2*	*MTRF1*
HDAC8	*IPO9*	*KIAA0922*	*LENG9*	*MDN1*	*MTUS1*
HDGFRP2	*IPP*	*KIAA1012*	*LEPREL1*	*MECP2*	*MUC12*
HDLBP	*IQCE*	*KIAA1109*	*LGR4*	*MED12*	*MUC16*
HEATR1	*IQGAP2*	*KIAA1211*	*LHFPL3*	*MED13*	*MUC4*
HEATR2	*IQGAP3*	*KIAA1244*	*LIAS*	*MED13L*	*MUC5AC*
HECTD1	*IQSEC1*	*KIAA1324L*	*LIG3*	*MED15*	*MUC5B*
HECW2	*IQSEC2*	*KIAA1432*	*LIMCH1*	*MED28*	*MUC7*
HEG1	*IRGC*	*KIAA1462*	*LINC00478*	*MEF2C*	*MYB*
HELLS	*ISLR2*	*KIAA1549*	*LINGO1*	*MEF2D*	*MYBBP1A*
HEPH	*ITGA11*	*KIAA1949*	*LIPE*	*MEGF11*	*MYCBP2*
HEXDC	*ITGA2*	*KIAA1967*	*LLGL1*	*MEGF8*	*MYCN*
HFE	*ITGA5*	*KIAA2018*	*LOXL4*	*MEIS2*	*MYH10*
HIGD2A	*ITGA6*	*KIAA2026*	*LPAR1*	*MEOX2*	*MYH13*
HIP1R	*ITGA9*	*KIDINS220*	*LPHN1*	*MEP1B*	*MYH2*
HIPK3	*ITGB2*	*KIF13A*	*LPIN1*	*MET*	*MYH6*
HIST1H2AE	*ITGB3*	*KIF18B*	*LPP*	*METTL14*	*MYH9*
HIST1H2AL	*ITGB4*	*KIF1A*	*LRBA*	*MFAP3*	*MYO15B*
HIST1H2BD	*ITGB6*	*KIF20B*	*LRCH3*	*MGAM*	*MYO1A*
HIST1H3A	*ITPR1*	*KIF5B*	*LRIG3*	*MGAT5*	*MYO1E*
HIST1H4C	*ITSN2*	*KIF5C*	*LRP1*	*MGAT5B*	*MYO1F*
HIVEP2	*IZUMO3*	*KIRREL*	*LRP2*	*MIA2*	*MYO1G*
HLTF	*JMJD1C*	*KIRREL3*	*LRP4*	*MIB1*	*MYO3A*
HMG20B	*JMJD7-PLA2G4B*	*KITLG*	*LRRC16A*	*MICALCL*	*MYO5A*
HMHA1	*JOSD1*	*KLB*	*LRRC31*	*MIOX*	*MYO5B*
HNRNPF	*JPH3*	*KLF12*	*LRRC4*	*MKI67*	*MYO7A*
HNRNPH1	*KALRN*	*KLHL10*	*LRRC8B*	*MKL2*	*MYO7B*
HNRNPU	*KANK1*	*KLHL11*	*LRRC8D*	*MKLN1*	*MYOF*
HOMER1	*KANSL1*	*KLHL15*	*LRRFIP2*	*MLH1*	*MYOM2*
HOOK	*KAT6A*	*KLHL18*	*LRRIQ*	*MLL*	*MYOM3*

MYSM1	*NOP2*	*OSBPL5*	*PHF17*	*PPP2R1A*	*RALGAPA2*
MYT1L	*NOP9*	*OSBPL9*	*PHF19*	*PPP2R1B*	*RALGAPB*
N4BP1	*NOS1*	*OTOGL*	*PHF20L1*	*PPP2R2B*	*RALGPS1*
N6AMT1	*NOTCH2*	*OVCH1*	*PHF3*	*PPP2R3A*	*RANBP17*
NAA10	*NOTCH3*	*P4HA3*	*PHIP*	*PPP2R5D*	*RANGAP1*
NAA15	*NOTUM*	*PABPC1*	*PHTF2*	*PPP3CB*	*RAPGEF1*
NAA40	*NPAP1*	*PACS1*	*PI4K2B*	*PPP6R1*	*RAPGEF6*
NAB2	*NPAS1*	*PACS2*	*PIAS1*	*PRCP*	*RARS*
NACA	*NPAS3*	*PAEP*	*PIGR*	*PRDM10*	*RASEF*
NAMPT	*NPFFR2*	*PAIP1*	*PIK3AP1*	*PRDM12*	*RASGRP1*
NANS	*NPLOC4*	*PAK3*	*PIK3C2B*	*PRDM4*	*RASIP1*
NAPRT1	*NR1H2*	*PAK6*	*PIK3R3*	*PRKCA*	*RASSF4*
NAV2	*NR2F1*	*PALB2*	*PION*	*PRKD1*	*RASSF5*
NBEA	*NR4A2*	*PALLD*	*PITPNM2*	*PRKDC*	*RB1*
NBEAL1	*NR5A2*	*PAMR1*	*PITPNM3*	*PRKX*	*RBBP4*
NBEAL2	*NR6A1*	*PAOX*	*PIWIL4*	*PRMT2*	*RBBP6*
NBN	*NRARP*	*PAQR4*	*PKD1*	*PRODH2*	*RBL2*
NCAPD2	*NRIP1*	*PAQR8*	*PKD1L1*	*PROM2*	*RBM12*
NCAPD3	*NRP1*	*PARN*	*PKD1L3*	*PROX2*	*RBM20*
NCAPG	*NRXN2*	*PARP15*	*PKD2L1*	*PRPF31*	*RBM23*
NCKAP1L	*NRXN3*	*PARP9*	*PKHD1*	*PRPF39*	*RBM33*
NCKAP5L	*NSD1*	*PASK*	*PKHD1L1*	*PRPF8*	*RBM47*
NCL	*NSUN7*	*PAWR*	*PKN2*	*PRPH2*	*RBMS1*
NCOR1	*NTM*	*PAX2*	*PLA1A*	*PRR19*	*RBX1*
NCOR2	*NTNG1*	*PAX3*	*PLA2G15*	*PRRC2B*	*RD3*
NDST4	*NTRK3*	*PBRM1*	*PLA2R1*	*PRUNE2*	*RDH8*
NDUFA10	*NTSR2*	*PCDH10*	*PLCB4*	*PSAT1*	*REEP5*
NDUFA9	*NUP107*	*PCDHA5*	*PLCD4*	*PSD2*	*REL*
NDUFB4	*NUP133*	*PCDHB1*	*PLCG2*	*PSEN1*	*RELN*
NDUFV1	*NUP98*	*PCDHB12*	*PLCH2*	*PSG7*	*RERGL*
NEB	*NYNRIN*	*PCDHB13*	*PLD3*	*PSMA7*	*REST*
NEBL	*OAZ2*	*PCDHB16*	*PLEC*	*PSMD3*	*REV3L*
NEDD4L	*OBSCN*	*PCDHB4*	*PLEKHA6*	*PSMG4*	*RFNG*
NEK1	*OBSL1*	*PCDHGA1*	*PLEKHA7*	*PTCH1*	*RFX2*
NEK11	*OCRL*	*PCDHGA12*	*PLEKHA8*	*PTCHD1*	*RFX3*
NEK5	*ODC1*	*PCDHGA4*	*PLEKHB2*	*PTCHD4*	*RFX7*
NEXN	*ODF2L*	*PCDHGB3*	*PLEKHN1*	*PTEN*	*RFX8*
NF1	*OGT*	*PCF11*	*PLG*	*PTGES*	*RGMA*
NFASC	*OPALIN*	*PCGF2*	*PLOD3*	*PTGIR*	*RGS14*
NFAT5	*OPCML*	*PCNX*	*PLXDC1*	*PTGR1*	*RGS22*
NFE2L1	*OPLAH*	*PDCD1*	*PLXNA1*	*PTK2B*	*RGSL1*
NGEF	*OPRL1*	*PDCD11*	*PLXNB1*	*PTK7*	*RHBDF1*
NHSL1	*OR10G4*	*PDCL2*	*PLXNB2*	*PTPN11*	*RHOBTB2*
NID2	*OR10H1*	*PDE11A*	*PLXND1*	*PTPN7*	*RHOG*
NIM1	*OR10K1*	*PDE4DIP*	*PML*	*PTPRA*	*RIC8B*
NIN	*OR10Q1*	*PDHA1*	*PNMAL1*	*PTPRK*	*RICTOR*
NIPA1	*OR10S1*	*PDIA3*	*PNPLA7*	*PTPRR*	*RIMS2*
NIPBL	*OR10Z1*	*PDIA5*	*PODXL2*	*PTPRS*	*RIOK3*
NIPSNAP1	*OR11L1*	*PDIA6*	*POGZ*	*PTPRT*	*RLBP1*
NISCH	*OR2C1*	*PDIK1L*	*POLA1*	*PUM2*	*RNF123*
NKTR	*OR2D2*	*PDS5B*	*POLQ*	*PURA*	*RNF145*
NKX2-1	*OR2L3*	*PEBP4*	*POLR2A*	*PWWP2A*	*RNF160*
NKX2-2	*OR2T1*	*PECR*	*POLRMT*	*PYCR2*	*RNF220*
NLGN1	*OR4X2*	*PEG3*	*POMP*	*QPRT*	*RNF26*
NLGN2	*OR5AC2*	*PER1*	*POPDC2*	*QRSL1*	*RNF38*
NLRP10	*OR5B3*	*PEX19*	*POR*	*R3HDM1*	*RNF41*
NLRP11	*OR5F1*	*PEX6*	*POU3F3*	*RAB2A*	*RNH1*
NLRP5	*OR5M1*	*PEX7*	*PPM1D*	*RAB31*	*ROBO1*
NLRP8	*OR6N2*	*PFKP*	*PPOX*	*RAB5C*	*ROBO2*
NLRX1	*OR6Y1*	*PGD*	*PPP1CB*	*RABGAP1L*	*ROBO3*
NOBOX	*OR7G2*	*PGM1*	*PPP1R13L*	*RABGGTB*	*ROCK1*
NOC3L	*OR8H1*	*PGM3*	*PPP1R15B*	*RAC1*	*ROS1*
No CCDS	*ORC1*	*PGR*	*PPP1R3B*	*RAD51*	*RP1*
NOLC1	*ORC3*	*PHACTR1*	*PPP1R9A*	*RAD54L2*	*RP11-439E19.8*
NONO	*OSBPL2*	*PHF12*	*PPP1R9B*	*RAI1*	*RP11-830F9*

RP1L1	*SERPINB12*	*SLC5A12*	*SPIN3*	*TATDN2*	*TMEM8A*
RPGRIP1	*SERPIND1*	*SLC5A6*	*SPOCD1*	*TBC1D1*	*TMPRSS12*
RPH3A	*SESN2*	*SLC6A1*	*SPOCK1*	*TBC1D17*	*TMPRSS15*
RPL34	*SET*	*SLC6A13*	*SPP2*	*TBC1D20*	*TMPRSS2*
RPRD1A	*SETBP1*	*SLC6A20*	*SPPL2B*	*TBC1D22B*	*TMPRSS5*
RPRD2	*SETD1B*	*SLC6A3*	*SPRR2D*	*TBC1D4*	*TMTC1*
RPS27A	*SETD2*	*SLC6A4*	*SPSB1*	*TBL1XR1*	*TNFAIP2*
RPS6KA3	*SETD4*	*SLC6A5*	*SPTAN1*	*TBL3*	*TNFRSF14*
RPTN	*SETD5*	*SLC6A8*	*SPTB*	*TBR1*	*TNK2*
RPTOR	*SETDB1*	*SLC6A9*	*SRBD1*	*TBRG1*	*TNKS*
RRP1B	*SETDB2*	*SLC7A7*	*SRCAP*	*TBX18*	*TNKS1BP1*
RRP8	*SF1*	*SLC8A3*	*SREBF1*	*TBX4*	*TNKS2*
RSF1	*SF3A3*	*SLC9A1*	*SREBF2*	*TCEB3B*	*TNN*
RSL1D1	*SFPQ*	*SLC9A3*	*SRGAP1*	*TCERG1L*	*TNPO2*
RSPH6A	*SGCE*	*SLCO1A2*	*SRGAP2*	*TCF20*	*TNPO3*
RTF1	*SGK2*	*SLCO1B7*	*SRGAP3*	*TCF3*	*TNRC6A*
RTKN2	*SGK223*	*SLCO1C1*	*SRPR*	*TCF4*	*TNXB*
RTN4RL1	*SGPP1*	*SLFN5*	*SRRM1*	*TCF7L1*	*TOB1*
RTN4RL2	*SGSM3*	*SMAD4*	*SRRM2*	*TCF7L2*	*TOP1*
RTP1	*SGTB*	*SMAD6*	*SRRM4*	*TCP10L2*	*TOP2B*
RTTN	*SH2D3C*	*SMARCA2*	*SS18L1*	*TCTN3*	*TOPORS*
RUVBL1	*SH3D19*	*SMARCA4*	*SSH2*	*TDRD1*	*TOX*
RUVBL2	*SHANK2*	*SMARCB1*	*SSPN*	*TDRD15*	*TP53INP2*
RXFP1	*SHC4*	*SMARCC1*	*ST3GAL3*	*TDRD5*	*TPK1*
RYBP	*SHKBP1*	*SMARCC2*	*ST3GAL6*	*TEAD4*	*TPM2*
RYR1	*SIGLEC9*	*SMARCD1*	*STAG1*	*TECTA*	*TPRG1*
RYR2	*SIK3*	*SMC1A*	*STAG2*	*TEKT3*	*TPTE2*
S1PR4	*SIN3A*	*SMC3*	*STARD9*	*TEKT5*	*TRAF6*
SAFB	*SIN3B*	*SMCHD1*	*STAT2*	*TEPP*	*TRAF7*
SAFB2	*SIRPB1*	*SMEK1*	*STAT4*	*TERT*	*TRAK1*
SAMD9	*SIX1*	*SMG9*	*STIL*	*TET1*	*TREH*
SAP130	*SKA3*	*SMTN*	*STIM2*	*TEX15*	*TRIM29*
SATB2	*SKAP1*	*SMTNL2*	*STK24*	*TFE3*	*TRIM32*
SBF1	*SKIL*	*SMU1*	*STK36*	*TFEC*	*TRIM5*
SBF2	*SLAMF1*	*SMURF1*	*STX1B*	*TFRC*	*TRIM6*
SCAF4	*SLAMF9*	*SNAP25*	*STXBP1*	*TGM6*	*TRIM8*
*SCAF4**	*SLC10A7*	*SNAPC3*	*STXBP3*	*THAP4*	*TRIM9*
SCAND3	*SLC13A1*	*SNAPC5*	*STXBP5*	*THOC2*	*TRIO*
SCAP	*SLC16A3*	*SNPH*	*SUCLA2*	*TIFA*	*TRIOBP*
SCARB2	*SLC17A3*	*SNRK*	*SULT1C4*	*TIMM44*	*TRIP10*
SCARF2	*SLC17A6*	*SNRNP200*	*SUPT16H*	*TIPIN*	*TRIP12*
SCFD2	*SLC1A2*	*SNRPA*	*SUPT6H*	*TJP2*	*TRMT61A*
SCHIP1	*SLC22A9*	*SNTG1*	*SUSD4*	*TJP3*	*TROAP*
SCN11A	*SLC24A3*	*SNX11*	*SUV420H1*	*TKT*	*TRPC5*
SCN1A	*SLC26A11*	*SNX22*	*SV2A*	*TLE2*	*TRPM2*
SCN2A	*SLC26A5*	*SNX30*	*SV2B*	*TLE3*	*TRPM5*
SCN3A	*SLC26A8*	*SNX9*	*SVEP1*	*TLK2*	*TRPM7*
SCN8A	*SLC26A9*	*SOGA1*	*SVIL*	*TLR1*	*TRRAP*
SCNN1D	*SLC27A2*	*SOGA3*	*SVOPL*	*TM7SF3*	*TSC2*
SCP2	*SLC2A1*	*SON*	*SYN2*	*TMC4*	*TSEN2*
SCRIB	*SLC30A10*	*SORCS1*	*SYNCRIP*	*TMCO5A*	*TSKS*
SCYL1	*SLC30A3*	*SORL1*	*SYNE2*	*TMEM105*	*TSNARE1*
SDCBP2	*SLC30A5*	*SOX2*	*SYNGAP1*	*TMEM154*	*TSNAXIP1*
SDK1	*SLC34A3*	*SOX2-OT*	*SYNJ1*	*TMEM205*	*TSPAN12*
SDK2	*SLC35A2*	*SOX5*	*SYNRG*	*TMEM214*	*TSPAN17*
SEC14L5	*SLC35F5*	*SP7*	*SYTL5*	*TMEM232*	*TSPYL1*
SEC16A	*SLC39A12*	*SPAG17*	*TAAR2*	*TMEM39B*	*TSPYL5*
SEC22C	*SLC39A2*	*SPARC*	*TAB1*	*TMEM42*	*TSR2*
SEC23IP	*SLC39A5*	*SPAST*	*TACC2*	*TMEM45A*	*TTBK1*
SEL1L	*SLC39A9*	*SPATA2*	*TAF1*	*TMEM56*	*TTC27*
SEMA3C	*SLC3A2*	*SPDEF*	*TAF8*	*TMEM61*	*TTC28*
SEMA4G	*SLC40A1*	*SPEN*	*TANC1*	*TMEM63B*	*TTC37*
SEMA7A	*SLC45A3*	*SPG7*	*TANC2*	*TMEM63C*	*TTC39A*
SENP5	*SLC5A10*	*SPHK2*	*TAS2R10*	*TMEM85*	*TTC39B*
SERAC1	*SLC5A11*	*SPIN1*	*TAS2R3*	*TMEM87B*	*TTC9C*

WAC	*WNT5A*	*ZBTB40*	*ZNF107*	*ZNF420*	*ZNF83*
WDFY2	*WNT8A*	*ZBTB41*	*ZNF12*	*ZNF425*	*ZNF831*
WDFY4	*WRN*	*ZC3H12B*	*ZNF133*	*ZNF451*	*ZNF839*
WDR19	*WWC1*	*ZC3HAV1*	*ZNF142*	*ZNF467*	*ZNF844*
WDR31	*WWOX*	*ZDBF2*	*ZNF146*	*ZNF479*	*ZNFX1*
WDR33	*XIRP1*	*ZDHHC1*	*ZNF155*	*ZNF498*	*ZRANB1*
WDR4	*XIRP2*	*ZEB2*	*ZNF175*	*ZNF516*	*ZRANB3*
WDR45	*XPO1*	*ZFC3H1*	*ZNF197*	*ZNF517*	*ZRSR2*
WDR60	*XPO5*	*ZFHX3*	*ZNF207*	*ZNF540*	*ZSCAN12*
WDR62	*XYLT1*	*ZFYVE16*	*ZNF213*	*ZNF541*	*ZSCAN2*
WDR65	*YIF1A*	*ZFYVE28*	*ZNF215*	*ZNF550*	*ZSCAN21*
WDR66	*YME1L1*	*ZFYVE9*	*ZNF217*	*ZNF638*	*ZWINT*
WDR73	*YTHDC2*	*ZIM2*	*ZNF225*	*ZNF642*	*ZZZ3*
WDR82	*YWHAG*	*ZKSCAN5*	*ZNF238*	*ZNF644*	
WDR87	*ZBED2*	*ZMYM2*	*ZNF292*	*ZNF700*	
WEE1	*ZBED4*	*ZMYM4*	*ZNF311*	*ZNF740*	
WHSC1L1	*ZBTB1*	*ZMYM6*	*ZNF335*	*ZNF761*	
WIZ	*ZBTB20*	*ZMYND11*	*ZNF341*	*ZNF778*	
WNK1	*ZBTB20-AS1*	*ZMYND12*	*ZNF362*	*ZNF783*	
WNK2	*ZBTB37*	*ZMYND8*	*ZNF365*	*ZNF813*	

Ion channel gene set

ANK2	*CLCN2*	*GRIN2A*	*KCNC2*	*KCNK1*	
BSND	*CLCN3*	*GRIN2B*	*KCNC3*	*KCNK10*	
KCNIP1	*CLCN4*	*GRIN2C*	*KCNC4*	*KCNK12*	
MINK1	*CLCN5*	*GRIN2D*	*KCND1*	*KCNK13*	
PSMD1	*CLCN6*	*GRIN3A*	*KCND2*	*KCNK15*	
SLC12A5	*CLCN7*	*GRINA*	*KCND3*	*KCNK16*	
TNRC15	*CLCNKA*	*GRM1*	*KCNE1*	*KCNK17*	
CACNA1A	*CLCNKB*	*GRM2*	*KCNE1L*	*KCNK2*	
CACNA1B	*DRD1*	*GRM3*	*KCNE2*	*KCNK3*	
CACNA1C	*DRD2*	*GRM4*	*KCNE3*	*KCNK4*	
CACNA1D	*DRD3*	*GRM5*	*KCNE4*	*KCNK5*	
CACNA1E	*DRD4*	*GRM6*	*KCNF1*	*KCNK6*	
CACNA1F	*DRD5*	*GRM7*	*KCNG1*	*KCNK7*	
CACNA1G	*GABBR1*	*GRM8*	*KCNG2*	*KCNK9*	
CACNA1H	*GABBR2*	*HCN1*	*KCNG3*	*KCNMA1*	
CACNA1I	*GABRA1*	*HCN2*	*KCNG4*	*KCNMB2*	
CACNA1S	*GABRA2*	*HCN3*	*KCNH1*	*KCNMB3*	
CACNA2D1	*GABRA3*	*HCN4*	*KCNH2*	*KCNMB4*	
CACNA2D2	*GABRA4*	*HTR1A*	*KCNH3*	*KCNN1*	
CACNA2D3	*GABRA5*	*HTR1B*	*KCNH4*	*KCNN2*	
CACNA2D4	*GABRA6*	*HTR1D*	*KCNH5*	*KCNN3*	
CACNB1	*GABRB1*	*HTR1E*	*KCNH6*	*KCNN4*	
CACNB2	*GABRB2*	*HTR1F*	*KCNH7*	*RYR1*	
CACNB3	*GABRB3*	*HTR2A*	*KCNH8*	*RYR2*	
CACNB4	*GABRD*	*HTR2C*	*KCNQ1*	*RYR3*	
CACNG1	*GABRE*	*HTR3A*	*KCNQ2*	*SCN10A*	
CACNG2	*GABRG1*	*HTR3B*	*KCNQ3*	*SCN11A*	
CACNG3	*GABRG2*	*HTR3C*	*KCNQ4*	*SCN1A*	
CACNG4	*GABRG3*	*HTR3D*	*KCNQ5*	*SCN1B*	
CACNG5	*GABRP*	*HTR3E*	*KCNRG*	*SCN2A2*	
CACNG6	*GABRQ*	*HTR4*	*KCNS1*	*SCN2B*	
CACNG7	*GABRR1*	*HTR5A*	*KCNS2*	*SCN3A*	
CACNG8	*GABRR2*	*HTR6*	*KCNS3*	*SCN3B*	
CHRNA1	*Gcom1*	*HTR7*	*KCNT1*	*SCN4A*	
CHRNA10	*GLRA1*	*KCNA1*	*KCNV1*	*SCN4B*	
CHRNA2	*GLRA2*	*KCNA10*	*KCNV2*	*SCN5A*	
CHRNA3	*GLRA3*	*KCNA2*	*KCNJ1*	*SCN7A*	
CHRNA4	*GLRB*	*KCNA3*	*KCNJ10*	*SCN8A*	
CHRNA5	*GRIA1*	*KCNA4*	*KCNJ11*	*SCN9A*	
CHRNA6	*GRIA2*	*KCNA5*	*KCNJ12*		
CHRNA7	*GRIA3*	*KCNA6*	*KCNJ14*		
CHRNA9	*GRIA4*	*KCNA7*	*KCNJ15*		
CHRNB1	*GRID1*	*KCNAB1*	*KCNJ16*		
CHRNB2	*GRID2*	*KCNAB2*	*KCNJ2*		
CHRNB3	*GRIK1*	*KCNAB3*	*KCNJ3*		
CHRNB4	*GRIK2*	*KCNB1*	*KCNJ4*		
CHRND	*GRIK3*	*KCNAB3*	*KCNJ5*		
CHRNE	*GRIK4*	*KCNB1*	*KCNJ6*		
CHRNG	*GRIK5*	*KCNB2*	*KCNJ8*		
CLCN1	*GRIN1*	*KCNC1*	*KCNJ9*		

Monogenic disorders with epilepsy gene set

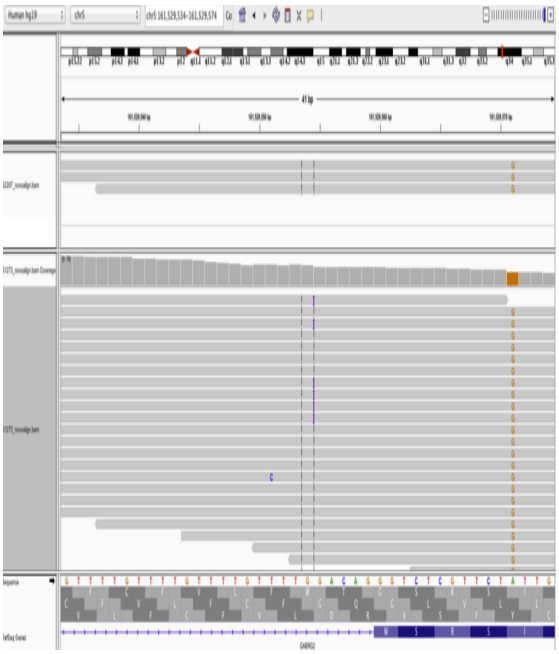
A2BP1	*CLCNKB*	*GLB1*	*NAGLU*	*RAB39B*	*TREX1*
ABCC8	*CLN3*	*GLDC*	*NDUFA2*	*RAB3GAP*	*TREX1*
ACY1	*CLN5*	*GLDC*	*NDUFS1*	*RAF*	*TSC1*
ADSL	*CLN6*	*GLI2*	*NDUFS3*	*RAI1*	*TSC2*
AGA	*CLN8*	*GLI3*	*NDUFS4*	*RARS2*	*TSEN2*
AHI1	*CNTNAP2*	*GLRA1*	*NDUFS7*	*RELN*	*TSEN34*
ALDH4A1	*COG1*	*GLRB*	*NDUFS8*	*RFT1*	*TSEN54*
ALDH5A1	*COG7*	*GNE*	*NDUFV1*	*RNASEH2A*	*TUBA1A*
ALDH7A1	*COG8*	*GNPTAB*	*NEU1*	*RNASEH2B*	*TUBA8*
ALG1	*COH1*	*GNPTG*	*NF1*	*RNASEH2C*	*TUBB2B*
ALG12	*COL18A1*	*GNS*	*NHEJ1*	*RPGRIP1L*	*UBE3A*
ALG2	*COL4A1*	*GPHN*	*NHLRC1*	*SAMHD1*	*VDAC1*
ALG3	*COQ2*	*GPR56*	*NIPBL*	*SCARB2*	*VDAC1*
ALG6	*COQ9*	*GPR98*	*NOTCH3*	*SCN1A*	*VPS13A*
ALG8	*COX10*	*GRIA3*	*NPC1*	*SCN1B*	*VRK1*
ALG9	*COX15*	*GRIN2A*	*NPC2*	*SCN2A*	*WDR62*
APTX	*CPT2*	*GRIN2B*	*NPHP1*	*SCN8A*	*ZEB2*
ARFGEF2	*CSTB*	*GUSB*	*NRAS*	*SCN9A*	*ZIC2*
ARG1	*CTSA*	*HEXA*	*NRXN1*	*SCO2*	
ARHGEF9	*CTSD*	*HEXB*	*OPHN1*	*SDHA*	
ARL13B	*CUL4B*	*HGSNAT*	*PAFAH1B1*	*SETBP1*	
ARSA	*CXORF5*	*HPD*	*PAK3*	*SGSH*	
ARSB	*DCX*	*HRAS*	*PANK2*	*SHH*	
ARX	*DCX*	*HSD17B10*	*PAX6*	*SHOC2*	
ASPA	*DLD*	*HYAL1*	*PCDH19*	*SIX3*	
ASPM	*DPAGT1*	*IDS*	*PCNT*	*SLC17A5*	
ATIC	*DPM1*	*IDUA*	*PC*	*SLC25A15*	
ATP1A2	*DPM3*	*INPP5E*	*PDHA1*	*SLC25A19*	
ATP2A2	*DPYD*	*JARID1C*	*PDSS1*	*SLC25A22*	
ATP6AP2	*EFHC1*	*KCNA1*	*PDSS2*	*SLC2A1*	
ATP6V0A2	*EIF2B1*	*KCNJ1*	*PEX1*	*SLC35A1*	
ATPAF2	*EIF2B2*	*KCNJ10*	*PEX12*	*SLC35C1*	
ATR	*EIF2B3*	*KCNMA1*	*PEX14*	*SLC46A1*	
ATRX	*EIF2B4*	*KCNQ2*	*PEX2*	*SLC4A10*	
B4GALT1	*EIF2B5*	*KCNQ3*	*PEX26*	*SLC6A5*	
BCS1L	*EMX2*	*KCTD7*	*PEX3*	*SLC9A6*	
BRAF	*EOMES*	*KIAA1279*	*PEX5*	*SMC1A*	
BRD2	*EPM2A*	*KRAS*	*PEX6*	*SMC3*	
BTD	*ETFA*	*L2HGDH*	*PEX7*	*SMPD1*	
BUB1B	*ETFB*	*LAMA2*	*PGK1*	*SMS*	
C12ORF65	*ETFDH*	*LARGE*	*PHF6*	*SNAP29*	
C8ORF38	*FGD1GPC3*	*LBR*	*PI12*	*SOS1*	
CABC1	*FGF8*	*LGI1*	*PIGV*	*SPRED1*	
CACNA1A	*FGFR3*	*LIG4*	*PLA2G6*	*SPTAN1*	
CACNA1A	*FH*	*LRPPRC*	*PLP1*	*SRPX2*	
CACNA1H	*FKRP*	*MAP2K1*	*PMM2*	*STIL*	
CACNB4	*FKTN*	*MAP2K2*	*PNKP*	*STXBP1*	
CASK	*FLNA*	*MAPK10*	*PNKP*	*SUMF1*	
CASR	*FOLR1*	*MBD5*	*POLG*	*SUMF1*	
CBL	*FOXG1*	*MCOLN1*	*POLGRARS2*	*SUOX*	
CC2D2A	*FUCA1*	*MCPH1*	*POMGNT1*	*SURF1*	
CCDC88C	*GABRA1*	*ME2*	*POMT1*	*SYNGAP1*	
CDK5RAP2	*GABRB3*	*MECP2*	*POMT2*	*SYP*	
CDKL5	*GABRD*	*MED17*	*PPT1*	*TACO1*	
CDKL5	*GABRG2*	*MFSD8*	*PQBP1*	*TBC1D24*	
CENPJ	*GALC*	*MGAT2*	*PQBP1*	*TBX1*	
CEP152	*GALNS*	*MLC1*	*PRICKLE1*	*TCF4*	
CEP290	*GAMT*	*MLL2*	*PRICKLE2*	*TGIF*	
CHRNA2	*GCDH*	*MOCS1*	*PRODH*	*TMEM15*	
CHRNA4	*GCS1*	*MOCS2*	*PSAP*	*TMEM216*	
CHRNB2	*GCSH*	*MPDU1*	*PTCH1*	*TMEM67*	
	GCST	*MPI*	*PTPN11*	*TMEM70*	
	GFAP	*MYST4*	*QDPR*	*TPP*	

Appendix I. IGV plots

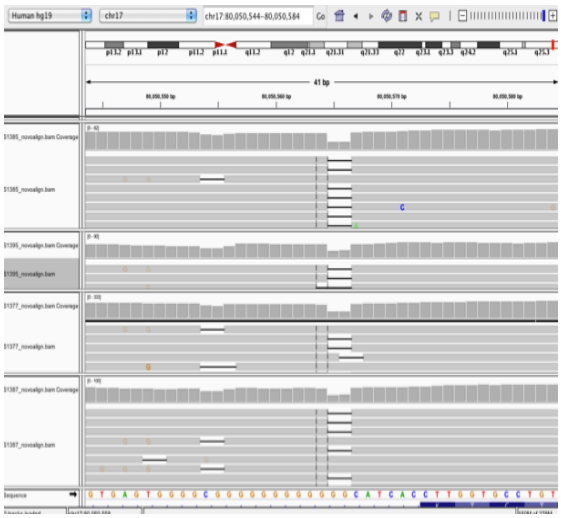
Novel epilepsy variants that failed IGV interrogation (see section 7.6)

S2294, ALDH7A1

S1273, GABRG2

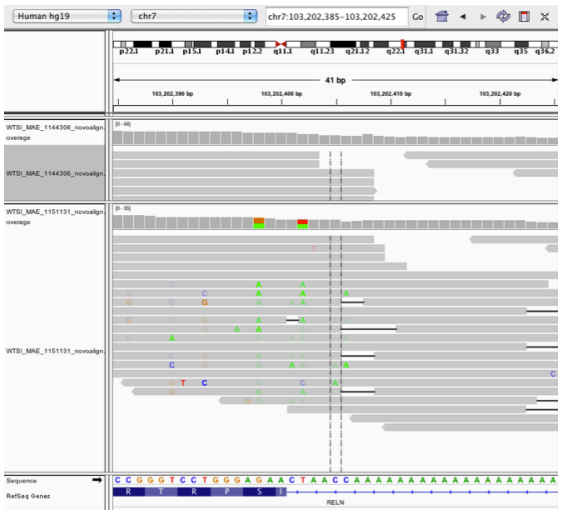


S1377, S1385, S1387, FASN

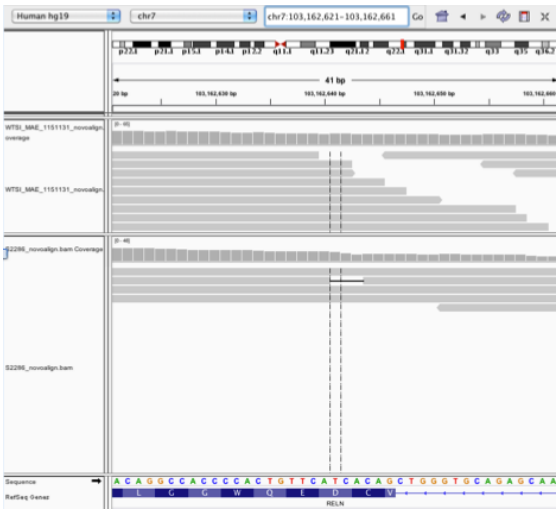


S2341, MECP2

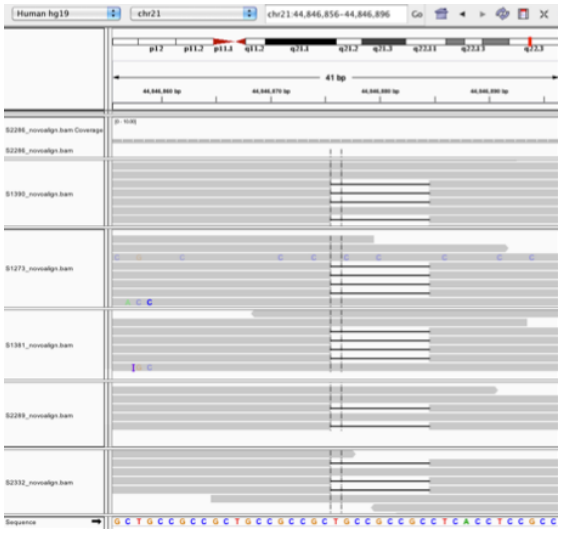
W_M_115113, RELN



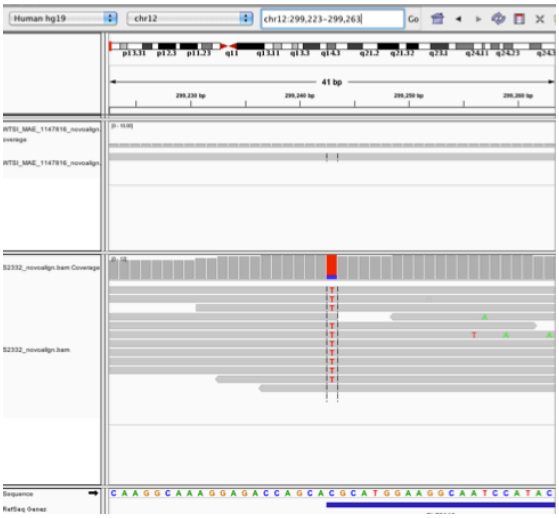
S2286, RELN



S1390, S1273, S1381, S2289, S2332, SIK1

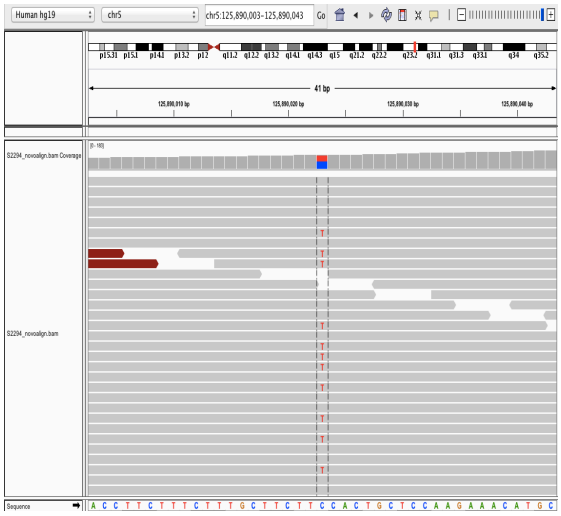


S2332, SLC6A1

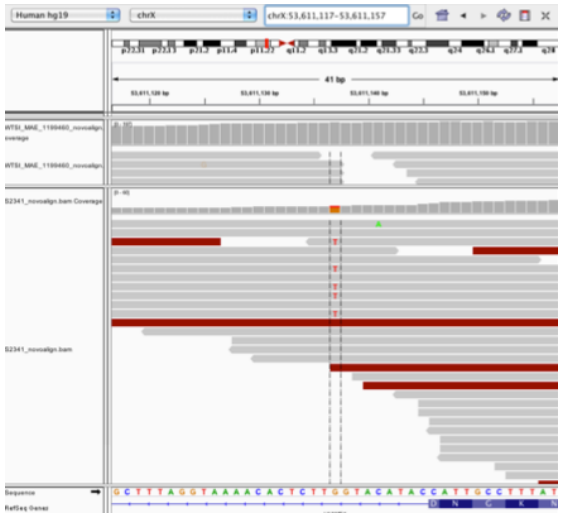


Likely benign variants (see section 7.6.1)

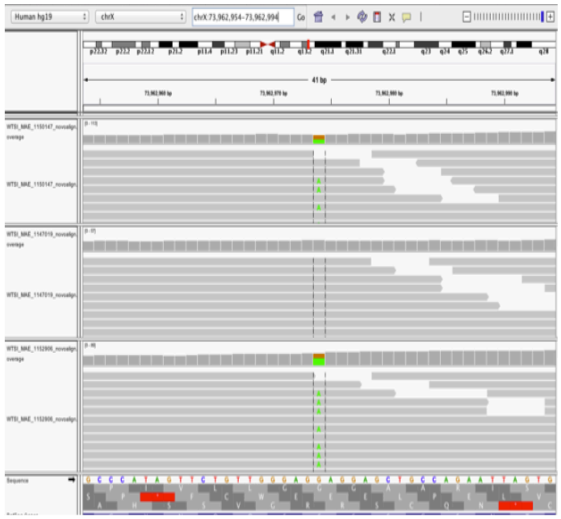
S2295, ALDH7A1



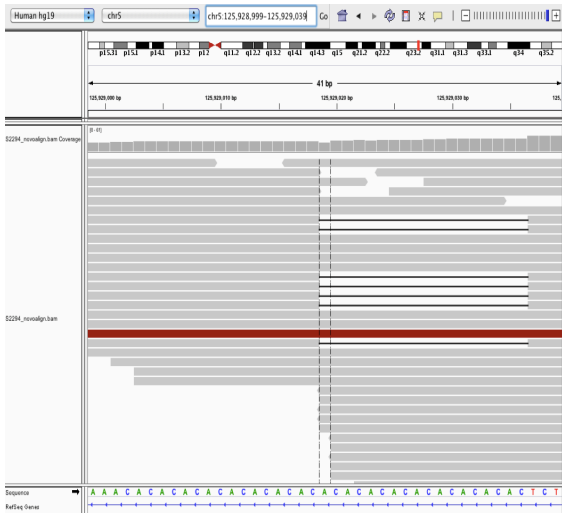
S2341, HUWE1



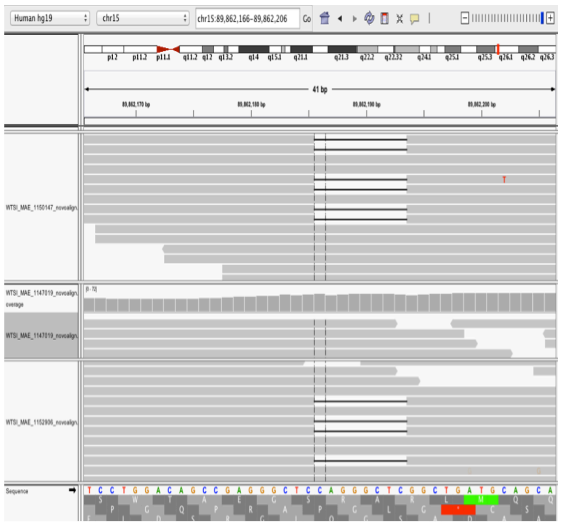
W_M_1150147, KIAA2022



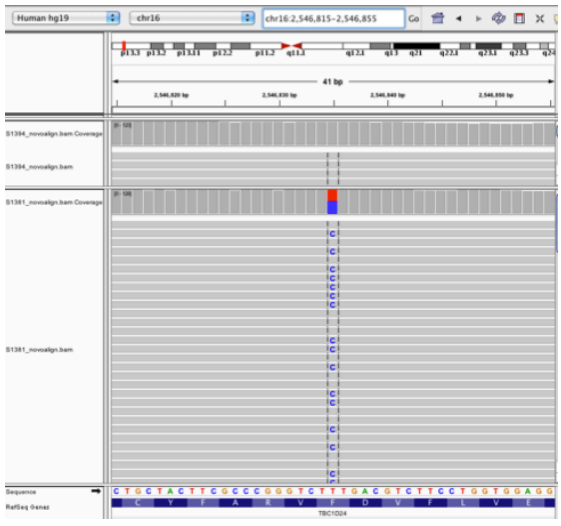
W_M_1195700, PLCB1



W_M_1150147, POLG

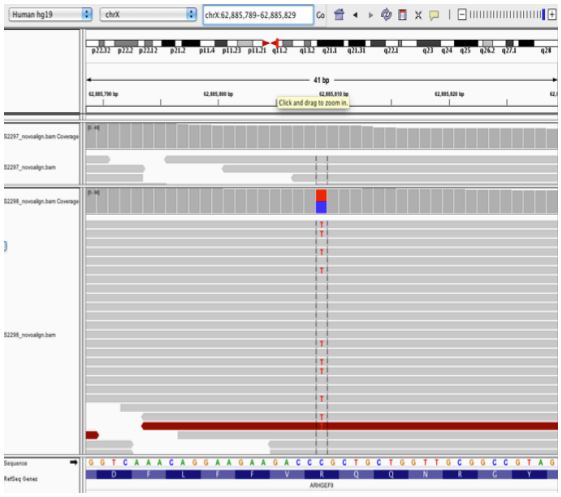


262N, TBC1D24

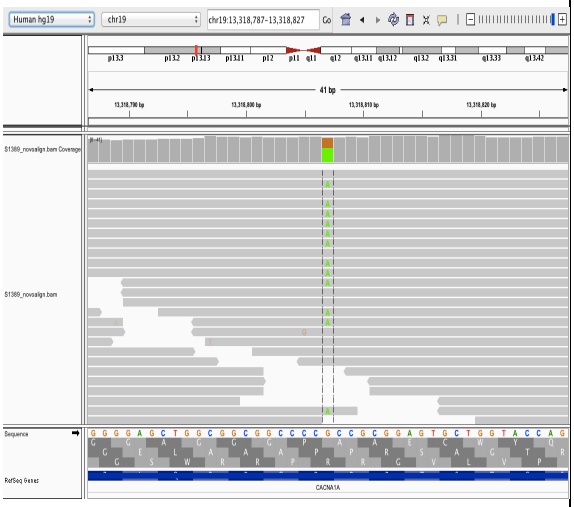


Variants of uncertain significance (from section 7.6.2)

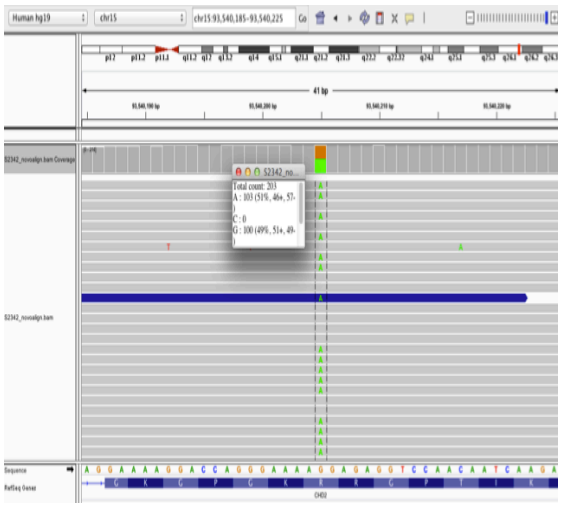
S2298, ARHGEF9



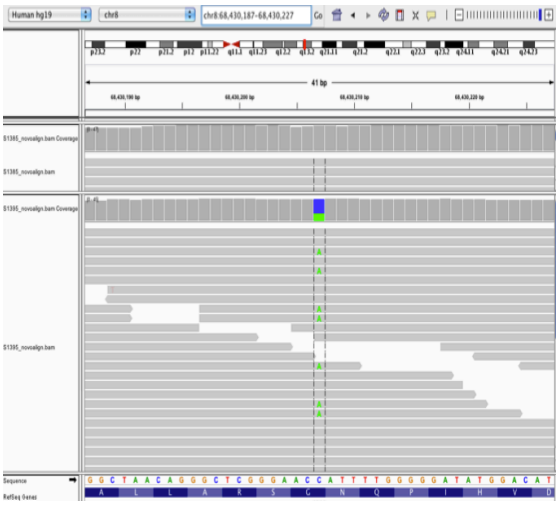
S1389, CACNA1A



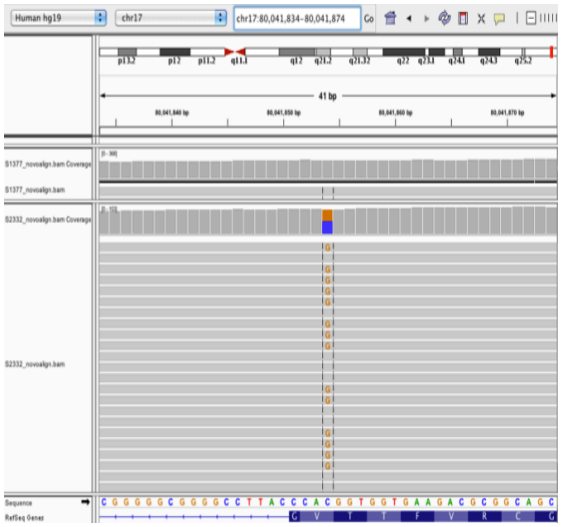
S2342, CHD2



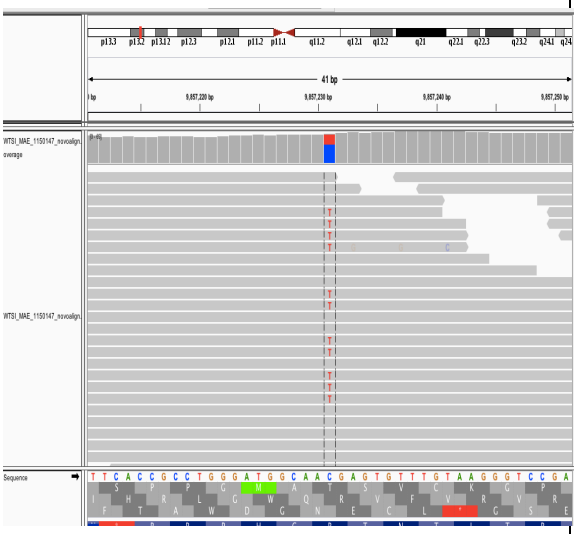
S1395, CPA6



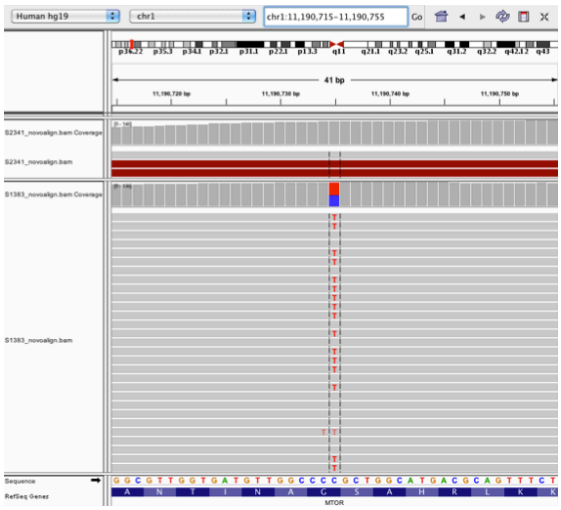
S2332, FASN



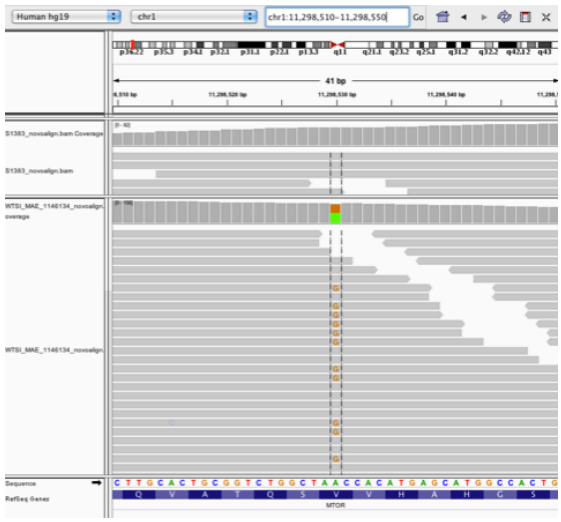
W_M_1150147, GRIN2A



S1383, MTOR

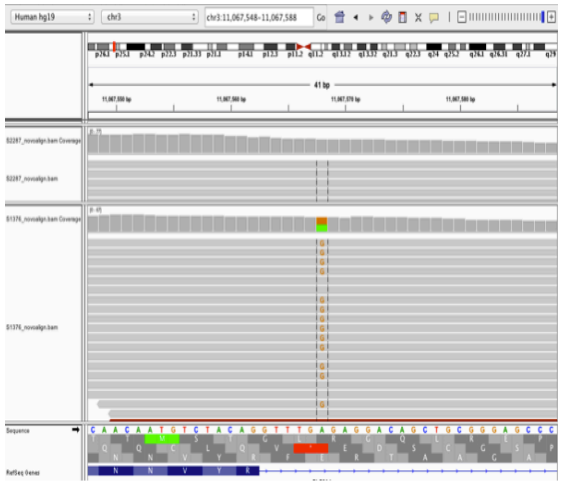


W_M_1146134, MTOR

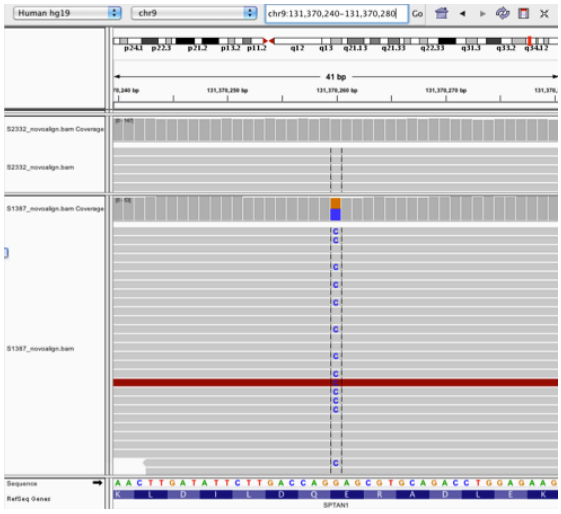


S2340, PIK3R2

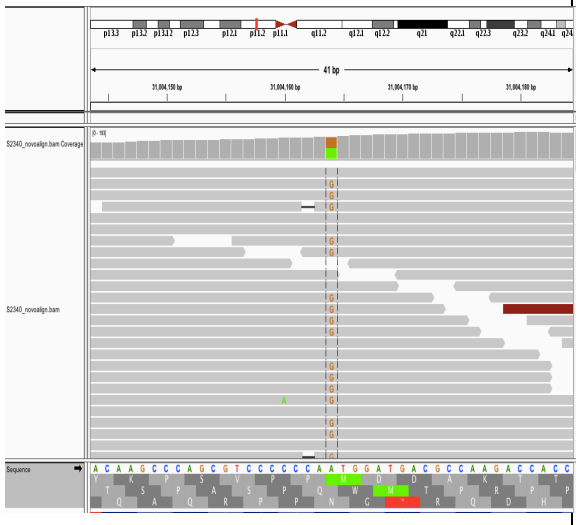
S1376, SLC6A1



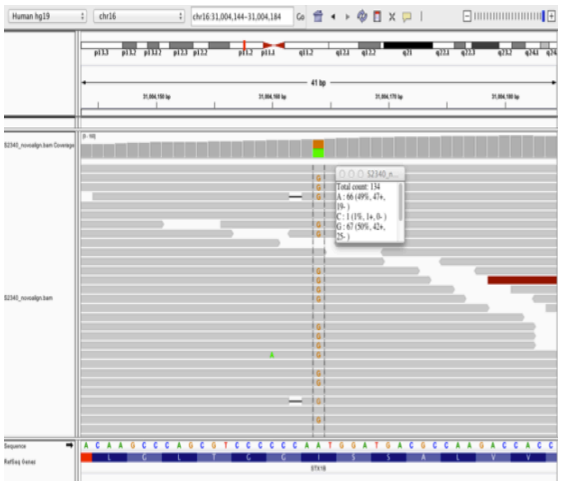
S1387, SPTAN1



W_M_1158865, STX1B

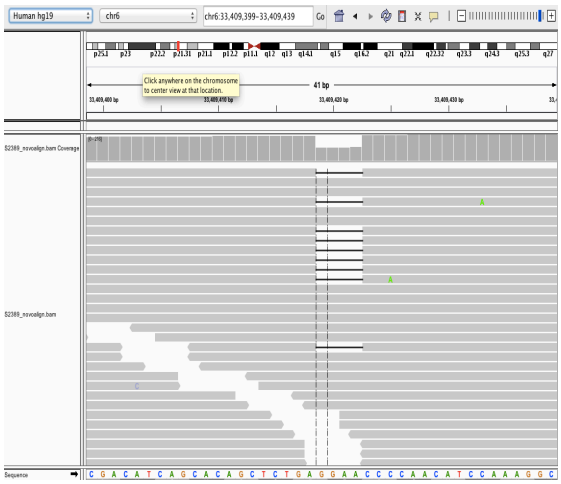


S2340, STX1B



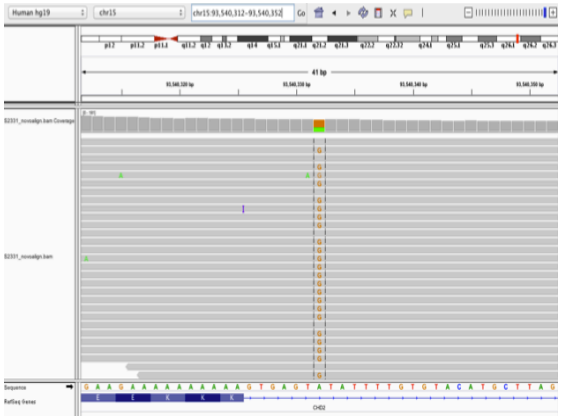
W_M_1150147, SYNGAP1

S2389, SYNGAP1

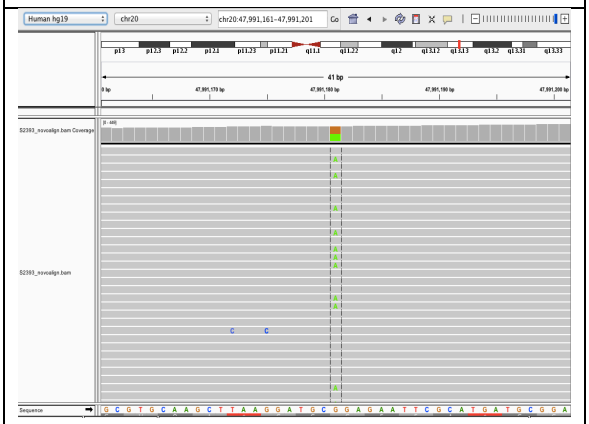


Likely pathogenic variants (from section 7.6.3)

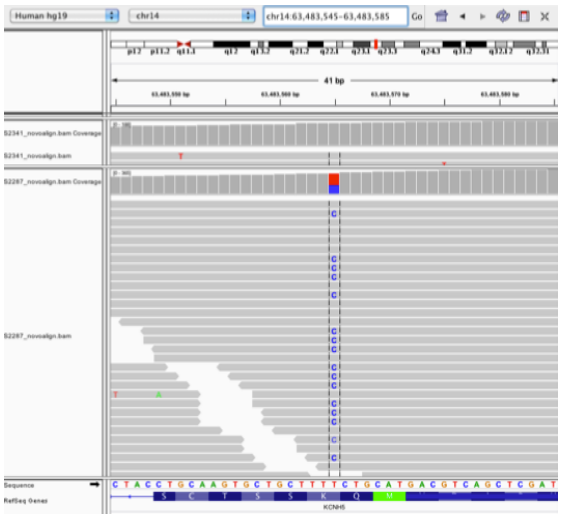
S2331, CHD2



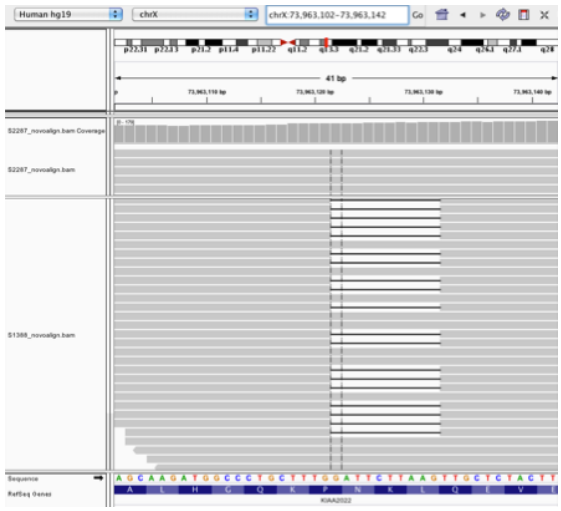
S2393, KCNB1



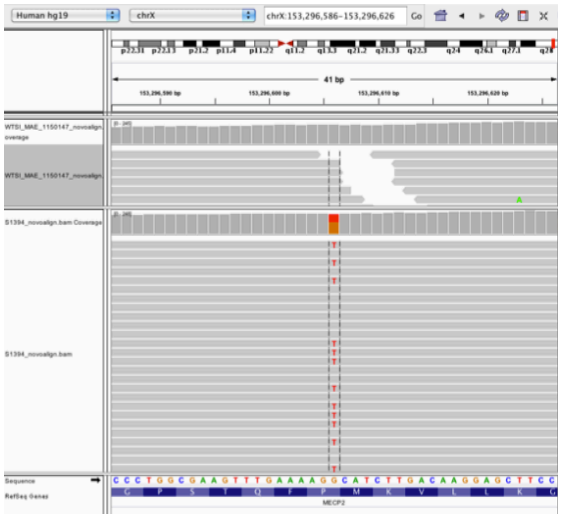
S2391, KCNH5



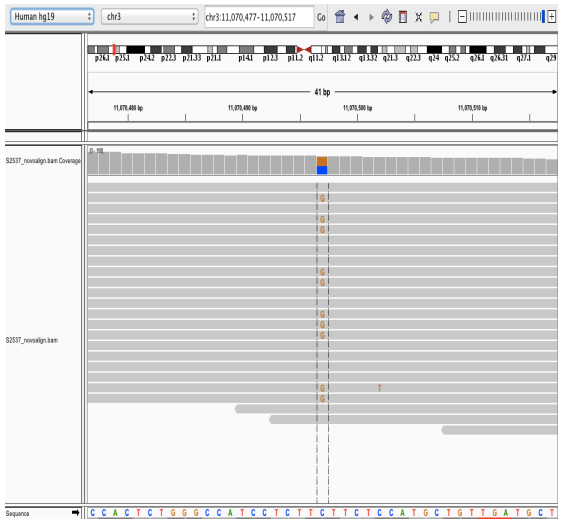
S1388, KIAA2022



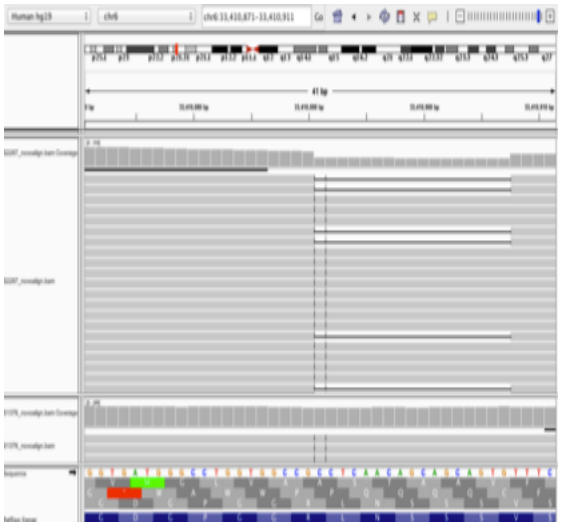
S1394, MECP2



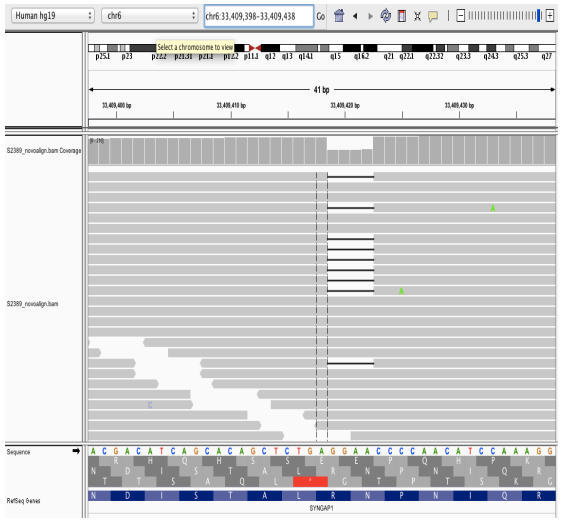
S2537, SLC6A1



S2287, SYNGAP1



S2389, SYNGAP1



Appendix J. Primers

Primer	Forward	Reverse	PCR
ASH1L_530_2FR	gggtgaaggagagtaaggca	tccccagcttcgaaatagaca	S
CHD2_546_1FR	aagaggtcagggttcgttga	TTTCTTCAGGGTCCACAGGG	S
CHD2_559_1FR	CCCTGTGGACCCTGAAGAAA	CAGCTTTGACACGACAGGTT	S
CLCN3_568_2FR	CCGTATTCCAAACCCACTTCC	ggtctgtcactgtaaaagggc	S
CLCN4_505_1FR	gcaacttcttcagcagcctt	ACAGTGGAAATTGTGGCTCC	Opt
KCNB1_533_1FR	TAGCTCCTCCGCAAAGTGAA	TGACTTGTTGGCCATTCTGC	S
KCNH5_525_1FR	GCAATACGTGACTGGAAAAGC	CCCAAAAGGCCTTAGTGAGC	Opt
KLHL1_518_1FR	cttgcttggcttcacaaacatc	TCTGCCTTACATTCACGACCTA	RT
KLHL1_518_2FR	TAAGAAGGGAGCTGGCCTTTAG	CCAATAACAACCAAACCAAACCA	RT
LOC4_533_1FR	GAACACGCCGGTCAGCAG	GGGTTTGACATAAGCGACTCC	S
LOC4_533_2F+R	GCTCTGGGGTCCACTGAG	CCACCTCCATTTTGTTAGGGT	S
Prickle2_530_1FR	GCTATGGTGCAATGAAGGGG	AGGAGAGTTCAGGGAGGACA	S
RYR3_505_1FR	ATCCTGTGTCTGGGTCGGAT	GGGTGTGGGGTGGAAGAAAA	S
SLC6A1_595_4FR	GGCTTAGAAAATGGGCAGTGT	TCTTCTTGGCAGCCTAAGAGT	Opt
SMARCA2_3003_1FR	tgccttgagaaatgggaccc	acagaaaaaggggcagaggt	S
STX1B_00504_1FR	TCTGTTTTGGGAGTGAGCCT	ACTGTCTGCCTGTCTCACTC	S
UBE3A_3000_1FR	GCCCAACTGTGTGCTTTCAT	TGTAGTTCTATGGTGGCCTCA	S

S Standard PCR, Opt Optimised PCR, RT real time PCR